CHAPTER 1

The Polymerase Chain Reaction

Getting Started

Charles R. M. Bangham

1. Introduction

The polymerase chain reaction (PCR) uses two oligonucleotide primers to direct the synthesis of specific sequences of DNA. One primer anneals to the coding strand of DNA and the other to the anticoding strand; the primer binding sites are typically separated by a few hundred base pairs (100–1000 bp). Repeated cycles of polymerization and denaturation lead to the exponential increase of the sequence defined by the primers. The extraordinary sensitivity and specificity of PCR have established it as a standard technique in molecular biology in the short time since it was first described *(1)*.

The purpose of this chapter is to suggest starting conditions for a PCR reaction and ways to overcome the main problems in PCR. It is intended as a practical guide, so theoretical aspects will not be discussed in detail. For a fuller account, there are excellent and comprehensive guides edited by Erlich *(2)* and by Innis et al. *(3)*. Protocols for special applications of PCR are described in later chapters in this volume.

2. Choice of Primers and Target DNA Sequence

The ideal oligonucleotide primer has the following features:

- Length: 18–30 bp. Shorter and longer primers may, however, work well. The primers should be similar in length and composition, so that their predicted melting temperatures (T_m, the temperature at which 50% of the strands are separated) are within 5°C.

From *Methods in Molecular Biology, Vol 9 Protocols in Human Molecular Genetics*
Edited by. C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

- GC content should be similar to the GC content of the template and of the other primer, ideally 50–60% GC.
- Binding site on target DNA: conserved region of sequence, ending on a nondegenerate base, e.g., first or second base of a conserved amino acid.
- No self-complementarity (to avoid secondary structures) or complementarity with the other primer. Computer programs are available to help identify such complementarity.
- No runs of three or more Gs or Cs at the 3' end of the primer.
- If mismatches between primer and template are known or likely to occur, these should be minimized at the 3' end of the primer, i.e., where the DNA polymerase binds. Highly degenerate primers may work under nonstringent reaction conditions, provided that at least three bases match at the 3' end of the primer *(4)*.
- Restriction sites can be included in the primer to help in efficient and directional cloning of the amplified product.

The ideal target sequence (template) to be amplified has the following features:

- Length: 150–500 bp. Lengths between 100 and 2000 bp can, however, often be amplified efficiently.
- Unique sequence, to avoid competition from unwanted templates.
- High copy number, to minimize the number of cycles of amplification required. PCR is, of course, highly efficient in detecting rare DNA species, but the risk of confusion with low-abundance contaminating DNA species increases if the target copy number is low.
- A diagnostic restriction enzyme site, to help verify amplification of the correct product.
- An intron sequence, to distinguish genomic amplification product from those amplified from cDNA or contaminating DNA (*see* Section 6).
- A sequence that can be detected specifically with a probe already in the laboratory.

3. Reagents

High-quality reagents are necessary for efficient amplification: particularly important are the DNA polymerase—usually the heat-stable enzyme from the thermophilic bacterium *Thermus aquaticus* (e.g., Perkin-Elmer/Cetus AmpliTaq®)—and the deoxynucleoside triphosphates (dNTPs). Stocks can be prepared as follows:

1. 1*M* KCl, 100 mL.
2. 1*M* Tris-HCl, pH 8.3 at 25°C, 100 mL.

3. 0.1 *M* $MgCl_2$, 100 mL.
4. 0.2% Gelatin (Difco), 100 mL.

Solutions 1–4 should be autoclaved and stored in 20-mL aliquots at room temperature.

5. Oligonucleotide primers: 50 μ*M* upstream primer (300 μg/mL of a 20-mer); 50 μ*M* downstream primer (300 μg/mL of a 20-mer).
6. 100 m*M* dNTPs at neutral pH (e.g., Pharmacia, Central Milton Keynes, Buckinghamshire, UK), stored at –80°C in aliquots of 5 or 10 μL.

To minimize the risk of cross-contamination with DNA templates from plasmids or previous amplification reactions, the stock solutions may be irradiated with UV light at 254 nm, e.g., 10 min in a Stratalinker 1800™ (Stratagene, Cambridge, UK). A 1-mL stock of 2× amplification solution containing all components except *Taq* polymerase and DNA template can be made, and stored at –20°C in 50-μL aliquots in siliconized 0.5-mL polypropylene tubes. The reaction mixture is then completed by adding the DNA template, *Taq* polymerase (e.g., 1–2.5 Cetus U of AmpliTaq®), and sufficient sterile water to bring the vol to 100 μL

4. Design of Reaction Mixture

For many purposes, the reaction mixture given in Table 1 will give efficient and specific amplification. However, there are a few variables that critically affect the efficiency and specificity of the reaction; the most important of these are the magnesium ion concentration and the oligonucleotide primer concentration (*see below* and Section 6).

The optimal number of DNA molecules in the template is between 10^5 and 10^6 *(3)*. For single-copy genes, this corresponds to approx 1 μg of human genomic DNA and 1 pg of a 6-kbp plasmid.

Optimization of the reaction mixture for a particular pair of oligonucleotide primers frequently involves two further steps:

1. Optimize Mg^{2+} concentration. Amplify the template with the following concentrations of Mg^{2+}: 1.5 (as above); 3.0; 4.5; 6.0; and 7.5 m*M*. Certain primer pairs may require further, finer adjustment of Mg^{2+} concentration, to within 0.5 m*M*.
2. Optimize primer concentrations. Amplify the template with the best Mg^{2+} concentration (as determined above), with the following concentrations of each primer: 0.05; 0.1; 0.25; 0.5; and 1.0 μ*M*.

Certain GC-rich templates do not amplify with the above protocol, probably because they rapidly adopt stable secondary structures on cooling from 94°C. The addition of dimethyl sulfoxide (DMSO) to the reaction mixture

Table 1
Basic PCR Reaction Mixture

Reagent	Final concentration, in 1×	Volume, µL, for 2× buffer, 50 µL	2× buffer, 1 mL
1*M* KCl	50 m*M*	5	100
1*M* Tris-HCl	10 m*M*	1	20
0.1*M* $MgCl_2$	1.5 m*M*	1.5	30
0.2% gelatin	0.01%	5	100
100 m*M* dGTP	200 n*M*	0.2	4
100 m*M* dATP	200 n*M*	0.2	4
100 m*M* dTTP	200 n*M*	0.2	4
100 m*M* dCTP	200 n*M*	0.2	4
50 µ*M* 5' primer	1 µ*M*	2	40
50 µ*M* 3' primer	1 µ*M*	2	40
Sterile, deionized water	–	33	654

(final concentration, 10%) may allow successful amplification, but this is not recommended in other cases, since it decreases the efficiency of the polymerase enzyme by about 50% *(5)*.

Addition of an overlay of inert mineral oil (about 50 µL) (e.g., paraffin oil BP, British Pharmacoepia) to the reaction mixture minimizes evaporation during amplification, and so increases the efficiency and reproducibility of the reaction *(6)*. However, it is not essential: if siliconized 0.5-mL tubes are used, the droplets that condense on the walls of the tube rapidly return to the solution. To reduce the number of components in the mixture, and so reduce the risk of DNA contamination, the mineral oil and gelatin, and in some instances the KCl, may be omitted.

5. Choice of Reaction Conditions

As with the reaction mixture design (Section 4), the following conditions serve to amplify efficiently and specifically in many cases. However, there are frequent instances in which the conditions need to be changed for a particular pair of primers. The most important variable to be optimized for a given primer pair is the annealing temperature. This adjustment is a highly empirical process; for example, the annealing temperature may need to be set at, or even above, the predicted T_m of a primer (note that the formula given below for estimating the T_m takes no account of the magnesium ion concentration).

5.1. Denaturation (94°C)

Incomplete denaturation is a frequent cause of failure of PCR. In the initial denaturation step, we use 5 min for a genomic DNA template and 2 min for a plasmid template. In subsequent cycles, 20–30 s at 94°C is adequate. If much longer times are required for successful amplification, the temperature in the reaction mixture itself should be measured with a thermocouple of low specific heat capacity to verify that the solution actually reaches the temperature required for denaturation.

5.2. Annealing (30–60 s)

First calculate the approximate T_m of the oligonucleotide primers, using a simple formula *(7)*, such as:

$$T_m = 2 \times (A + T) + 4 \times (G + C)$$

in °C.

Then set the annealing temperature at 5°C below the lower of the two predicted T_ms.

If nonspecific amplification products are a particular problem, annealing and extension can be performed in a single step at between 60 and 72°C.

5.3. Extension (72°C)

Allow 1 min/1 kbp of desired product. If the required product is short (<200 bp), then the extension step can be omitted, since *Taq* polymerase is sufficiently active at lower temperatures to complete the reaction during the transition between the annealing temperature and the denaturation temperature.

6. Troubleshooting in PCR

It is now widely realized that the remarkable sensitivity of PCR is also its main limitation, because a single contaminating molecule of DNA containing the target sequence may be amplified, leading to potentially serious misinterpretation of the results *(8–11)* The standards of cleanliness required in making up the solutions are therefore higher than for almost any other laboratory procedure, albeit for different reasons.

The main precautions to be taken to avoid false positive results in PCR are listed in Table 2, in approximate order of importance. It is essential to include in each experiment a tube containing all the components except the DNA template, and to examine the products on a gel stained with ethidium bromide, to look for contamination of the reaction mixture.

Table 2
Precautions to Avoid DNA Contamination of PCR Reactions

1. Make up reaction mixtures in a laboratory in which plasmids containing the target sequence are never handled. *Never* take amplified product into this laboratory. Many workers make up their PCR solutions in laminar-flow hoods.
2. Wear gloves when making up solutions; avoid touching the inside of the tube cap.
3. To make up reaction mixtures, use pipets that are never used to handle plasmid or amplification products with the appropriate sequence. We recommend "positive displacement" pipets (e.g., Gilson "Microman"), with tips containing disposable plungers that prevent aerosol contact or direct contact between the pipet barrel and the solution. For handling small volumes (0.5–10 µL), calibrated disposable glass microcapillaries are very useful (e.g , Drummond PCR micropipets).
4. Aliquot reagents and reaction buffers, and use each aliquot only once. *See also* Sections 2 and 3.
5. Irradiate solutions used in PCR with UV. This was shown to abolish the amplification of plasmid that was deliberately added to PCR mixtures *(8)* Solutions containing all components except the DNA template can safely be irradiated for 10 min on a 305-nm-wavelength laboratory UV transilluminator, without denaturing the primers or the enzyme.
6. Avoid reamplification of primary amplified products, if possible. If amplification of gel-purified DNA fragments is necessary, irradiate the agarose gel and its running buffer in the gel apparatus with 254nm W before running: 10 min in a Stratalinker 1800™ (Stratagene) is sufficient.
7. Some workers find that contamination is abolished only when the person making up the solutions wears a surgical face mask and sometimes a hair net (*9*, J. Todd, personal communication).

In some cases (for example, in RNA viruses), it may be possible to amplify between conserved nucleotide sequences, across a highly variable sequence. If the frequency of nucleotide differences between two amplified products greatly exceeds the error rate of *Taq* polymerase, then DNA cross-contamination can be excluded beyond reasonable doubt *(11,12)*.

The dose of UV radiation required to prevent amplification depends on the size and the base composition of the potential contaminating species *(13)*. Ideally, the dose should be titrated with a given template and a known contaminant with the UV source used in the laboratory.

The other common problems in PCR relate to the specificity and efficiency of amplification of the required product, avoiding amplification from partial matches between the primers and template. Some of these have been addressed above (*see* Sections 3 and 5); a summary of the most frequent causes and their remedies is given in Table 3.

Table 3
Troubleshooting in PCR

Problem	Causes	Remedies
No detectable product after repeated attempts	Inadequate melting of DNA template	Increase time in denature step
	Target sequence too rare	Increase number of cycles (up to 60)
	Annealing temperature too high	Lower temperature by 5°C
	GC content of target sequence too high	Try 10% DMSO in reaction
	Primers anneal to each other (primer dimer) or to themselves	*See note* a
Multiple bands on agarose gel of amplified product	"Overamplification"[b]	Reduce number of cycles; reduce extension time
	Primers too short or degenerate	*See note* a
	Concentration of dNTPs or of enzyme too high	Reduce either by 2–10×
Continuous "smear" of amplified product on agarose gel	Annealing temperature too low for GC content of primers	Raise annealing temperature by 5°C
	"Overamplification"[b]	
Predominance of very high mol wt amplified product	Reamplification of primary amplified product	Reduce number of cycles Gel-purify primary product before reamplification
"Primer dimer"[c]	Complementarity between 3' ends of primers	*See note* a

[a] In each case, the remedy is to increase the stringency of the reaction by increasing the annealing temperature *or* reducing the primer concentration *or* both

[b] "Overamplification" denotes the use of too many cycles of PCR, which favors the amplification of mismatched or nonspecific DNA products. For amplification of a single-copy gene from genomic DNA, 35 cycles should be enough, but more cycles may be needed for a rare species, such as a low-copy-number infectious agent

[c] The "primer dimer" results from annealing and polymerization of the 5' primer on the 3' primer, and appears as a fuzzy low-mol-wt band on an ethidium bromide stained agarose gel

If the problem is one of persistent failure to amplify any band, it may be necessary to choose a different sequence for one or both primers: certain sequences are very inefficient as PCR primers, for unknown reasons. If this is suspected, each primer should be tested in a PCR reaction with another PCR primer of demonstrated efficacy, from the same template sequence (if available). In this way it is frequently possible to show which of the two primers is at fault.

References

1. Saiki, R., Scharf, S., Faloona, F., Mullis, K. B., Horn, G T , Erlich, H A , and Arnheim, N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia. *Science* **230,** 1350–1354
2. Erlich, H. A., ed. (1989) *PCR Technology: Principles and Application for DNA Amplification.* Stockton, New York.
3. Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J , eds (1990) *PCR Protocols. A Guide to Methods and Applications.* Academic, New York.
4. Sommer, R. and Tautz, D. (1989) Minimal homology requirements for PCR primers *Nucleic Acids Res.* **17,** 6749.
5. Gelfand, D. H. and White, T. J. (1990) Thermostable DNA polymerases, in *PCR Protocols: A Guide to Methods and Applications* Innis, M A , Gelfand, D H , Sninsky, J J , and White, T. J., eds. Academic, New York, p. 129
6. Mezei, L. M. (1990) Effect of oil overlay on PCR amplification, in *Amplifications* Perkin-Elmer, Norwalk, CT, vol. 4, p. 11.
7. Thein, S. L. and Wallace, R. B. (1986) The use of synthetic oligonucleotides as specific hybridization probes in the diagnosis of genetic disorders, in *Human Genetic Diseases: A Practical Approach.* K. E. Davies, ed. IRL, Oxford, UK, pp. 33–50
8. Sarkar G. and Sommer, S. S. (1990) Shedding light on PCR contamination. *Nature* **343,** 27.
9. Kitchin, P. A., Szotyori, Z., Fromholc, C , and Almond, N (1990) Avoidance of false positives. *Nature* **344,** 201.
10. Kwok, S. and Higuchi, R. (1989) Avoiding false positives with PCR *Nature* **339,** 237,238
11. Bangham, C. R. M , Nightingale, S., Cruickshank, J K., and Daenke, S. (1989) PCR analysis of DNA from multiple sclerosis patients for the presence of HTLV-I *Science* **246,** 821.
12. Daenke, S., Nightingale, S., Cruickshank, J. K., and Bangham, C R M (1990) Sequence variants of human T-cell lymphotropic virus type I from patients with tropical spastic paraparesis and adult T-cell leukemia do not distinguish neurological from leukemic isolates. *J. Virol.* **64,** 1278–1282
13. Cimino, G. D., Metchette, K., Isaacs, S. T., and Zhu, Y. S. (1990) More false positive problems. *Nature* **345,** 773,774.

CHAPTER 2

Direct DNA Sequencing of Complementary DNA Amplified by the Polymerase Chain Reaction

Richard A. Gibbs, Phi-Nga Nguyen, and C. Thomas Caskey

1. Introduction

Protocols for the sequence analysis of conventional single-stranded or double-stranded DNA templates are often unsuitable for the direct sequencing of DNA fragments generated by the polymerase chain reaction (PCR) *(1,2)*. The features that can distinguish PCR products as templates for sequencing include (a) contamination of the reactions by nonspecific PCR amplification products that are complementary to the sequencing primer, (b) the persistence of "leftover" PCR primers from the amplification reactions, and (c) the potential for competition between one strand of the amplified fragment and the oligonucleotide used for the sequencing. The various approaches that have been used to overcome these problems include

1. The use of 5'-end-labeled DNA-sequencing primers that are complementary to regions between the PCR primers *(3)*;
2. Gel purification of amplified DNA to remove unwanted fragments and primer *(4)*;
3. Spin columns for the separation of leftover primers from high mol wt material *(5, 6)*;
4. "Asymmetric" or "unbalanced" PCR priming to generate an excess of single strands during the initial amplification *(7)*;

From: *Methods in Molecular Biology, Vol. 9. Protocols in Human Molecular Genetics*
Edited by C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

5. Addition of dimethylsulfoxide (DMSO) to sequencing reactions with short annealing times *(8)*; and
6. The use of several short, high-temperature, sequencing cycles *(9)*.

In developing the protocol that is described here (summarized in Fig. 1), we have endeavored to avoid the tedious steps of gel purification or column chromatography. Instead, we have developed a two-step reaction procedure for template preparation that first allows amplification of a specific fragment and then the production of an excess of one strand. This method is essentially a modification of the asymmetric priming protocol of Gyllensten and Erlich (7). The current method can be performed comfortably in two days and enables the reliable generation of DNA sequence ladders that can be resolved as far as the gel system that is used will allow. The technique has been applied for the analysis of transcribed human sequences, for which it is preceded by a reverse transcription reaction. Equal success has been obtained in the analysis of human gene sequences using 10–100 ng of genomic DNA as starting material and there is no reason that virtually any DNA fragment that can be successfully amplified by PCR would not be amenable to this analysis. Features that are modifications of other protocols or that we regard as particularly important are further discussed below.

2. Materials (*see* Note 1)

1. Ribonuclease inhibitor (RNasin; Pharmacia, catalog no. 27-0815-01; 27,000 U/mL).
2. Random hexamer primers ($pd(N)_6$, Pharmacia, catalog no. 27-2166-01; at 5 mg/mL).
3. 5× POL buffer (250 m*M* Tris-HCl, pH 8.3 at 37°C, 40 m*M* $MgCl_2$, 150 m*M* KCl; 50 m*M* dithiothreitol [DTT]).
4. Deoxyribonucleotide triphosphate mixture (dNTPs; mixture of 25 m*M* each of dATP, dTTP, dCTP, and dGTP).
5. Reverse Transcriptase (M-MuLV, Pharmacia catalog no. 27-0925-02;12,000 U/mL).
6. Modified T7 DNA polymerase (Sequenase™; usually supplied at 12.5 U/µL from United States Biochemicals [USB] catalog no. 70722).
7. Dideoxynucleotide terminator mixtures; these are 80:8 µ*M* deoxy:dideoxy mixtures. The solutions supplied by USB, catalog nos. 70714 ("A" mix), 70716 ("C" mix), 70718 ("G" mix), and 70720 ("T" mix) are appropriate. The solutions are thawed and stored in 20-µL aliquots. Sequenase™ (1.0 µL) is added to each just before use.
8. Sequencing stop solution (STOP; 95% formamide, 20 m*M* EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol).

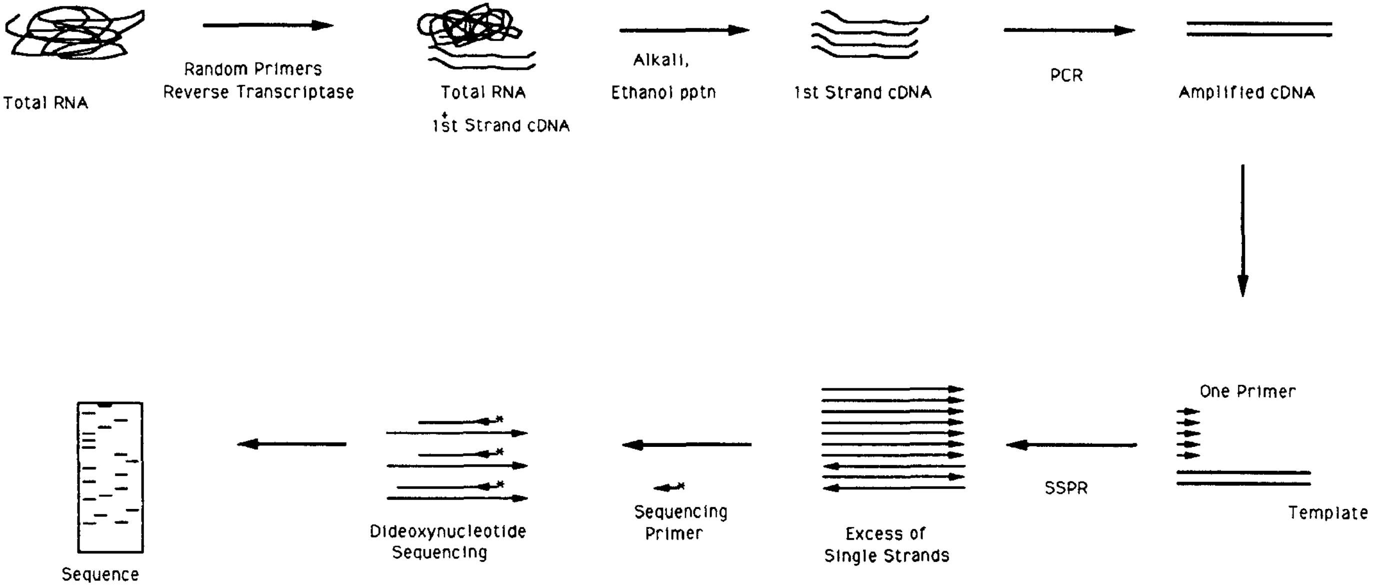


Fig. 1. Schematic diagram of the strategy for direct DNA sequencing of PCR-amplified cDNA. Total cellular RNA is copied by randomly primed reverse transcription, the RNA is hydrolyzed by alkali, and the specific cDNA is amplified by PCR. An aliquot of the PCR product is used in a single-strand-producing reaction (SSPR) that has a single oligonucleotide primer. The single-stranded mixture is sequenced using a 5'-end-labeled primer and dideoxynucleotide termination.

9. Reverse-transcription hydrolysis solution (RTH): 0.7*M* NaOH, 40 m*M* EDTA.
10. 2*M* ammonium acetate, pH 4.5.
11. 10× PCR buffer: 67 m*M* $MgCl_2$, 166 m*M* $(NH_4)_2SO_4$, 28 m*M* 2-mercaptoethanol, 68 µ*M* EDTA, 670 m*M* Tris-HCl; pH 8.8 at 25°C.
12. *Taq* DNA polymerase (Perkin-Elmer/Cetus).
13. 7.5*M* ammonium acetate.
14. [^{32}P] γ-TP (6000 Ci/mmol).
15. T4 polynucleotide kinase.
16. 10× kinase buffer: 100 m*M* Tris-HCl, pH 7.6; 100 m*M* $MgCl_2$; 100 m*M* DTT.
17. NENsorb® columns (New England Nuclear/Dupont).
18. TE buffer: 10 m*M* Tris-HCl, pH 7.0; 1 m*M* EDTA.

3. Methods

3.1. Reverse Transcription Reactions (see Notes 2,3)

1. Mix the following on ice: 0.5–5.0 µg total cellular RNA; 0.5 µL RNasin; 2.0 µL pd(N)$_6$ primers; 4 µL of 5× POL buffer; and H_2O (treated with diethylpyrocarbonate) to 15.5 µL.
2. Heat at 95°C for 1 min, chill on ice, and pulse/spin in a microfuge. Then add, at room temperature, 2.0 µL of dNTPs, 0.5 µL of RNasin, and 1.0 µL of reverse transcriptase.
3. Incubate at 37°C for 1 h.
4. Add 30 µL of RTH solution, mix gently, and incubate at 65°C for 10 min.
5. Add 5 µL of 2*M* ammonium acetate (pH 4.5), mix, add 130 µL of ethanol, and chill at –20°C for at least 4 h (preferably overnight). Then spin, wash in 70% ethanol, wash again in 100% ethanol, and dry.

3.2. Polymerase Chain Reaction (see Note 4)

1. Mix 5–10% of the product of one cDNA-synthesis reaction with 50 pmol of each PCR primer (*see* Note 9) in a total vol of 50 µL containing 5 µL of 10× PCR buffer, 1.5 m*M* of each dNTP (3 µL of 25 m*M* mixture) and 10% DMSO. (This buffer is a slight modification of that described by Kogan et al. *[10]*.)
2. Heat to 94°C for 5 min and centrifuge for 5 s.
3. Add 2.5 U of *Taq* DNA polymerase, mix gently, and overlay with mineral oil.
4. We typically perform 23–28 cycles of DNA polymerization (68°C, 1–3 min), denaturation (94°C, 30 s), and annealing (37–65°C, 30 s). The optimum annealing temperature must be determined empirically In initial reactions, allow at least 1 min of extension/500 bases.

5. The final incubation at 68°C is extended for 7 min.
6. Remove the sample from under the oil.

3.3. Single-Strand-Producing Reactions (SSPRs) (see Notes 5–8)

1. Take 1 µL of the PCR product to initiate a second PCR that is identical to the first except that only one primer is used. Use a primer that is opposite in sense to the sequencing primer that will be employed.
2. Perform the same number of cycles of the SSPR as was used for the initial PCR. Use the same cycling temperatures, but double the length of the annealing and polymerase extension times.
3. Dilute the reaction with an equal volume of H_2O and add an equal volume of 7.5*M* ammonium acetate, mix, add 2.5 vol of ethanol, chill for 15 min at –70°C (or overnight at 4°C), and spin for 30 min in a microfuge. Repeat the ammonium acetate precipitation. Wash with 70% ethanol, again with 100% ethanol, and dry to completion under vacuum. Dissolve the pellet in 10 µL of H_2O immediately before use in the DNA-sequencing reaction

3.4. Radiolabeling the DNA Sequencing Primer (see Note 9)

1. Kinase reactions contain 20–50 pmol of primer, 50–70 µCi of [^{32}P] γ-ATP (6000 Ci/mmol), 30 U of T4 polynucleotide kinase, and 5 µL of 10× kinase buffer in a total vol of 50 µL. Reactions proceed at 37°C for 45 min.
2. Purify the labeled primer by passage through a NENsorb® column. Dry the product to completion, and then resuspend in 12 µL of H_2O immediately before use.

3.5. DNA-Sequencing Reactions (see Notes 10–13)

1. Add 5.0 µL of single-strand DNA template to 3 µL of labeled primer and 2.0 µL of 5× POL buffer, in a standard 1.5-mL microcentrifuge tube.
2. Heat to 95°C for 10 min.
3. Centrifuge for 5 s to bring down condensation.
4. Dispense 2.5-µL aliquots of the primer template mixture into four appropriately labeled tubes (1T, 1C, 1G, 1A). Do this step on the bench, i.e., at room temperature.
5. Add 2.0 µL of the appropriate dideoxy-terminator/Sequenase™ mix *(see above)* to each of the four tubes and place immediately at 50°C. Incubate for 10 min.

6. Centrifuge for 5 s to bring down condensation and add 3.0 µL of STOP solution.
7. Heat to 80°C for 2 min and analyze by electrophoresis and autoradiography (*see* Chapter 3).

4. Notes

1. Recommended manufacturers: These recommendations are not meant to imply that only the specified manufacturer products can be used.
2. Synthesis of the first cDNA strand: Syntheses of cDNA have been performed from poly(A+) RNA, total cellular RNA or crude cell extracts *(11,12)*. We always prepare total cellular RNA by the guanidinium method *(13)*, which is convenient when a relatively small number of samples are to be analyzed. There is no need to prepare poly(A+) RNA, although if you already have some it works fine.

   There are at least three methods for priming, the synthesis of cDNA· random priming, oligo (dT) priming, and specific oligimer priming. Priming with a specific oligimer has been avoided, since the resulting cDNA cannot be used as a template for PCR of other DNA fragments. In addition, the conditions for annealing of a specific oligimer must be stringently controlled. There seems to be little difference in performance among the nonspecific priming methods, although oligo (dT) has the theoretical disadvantage of less efficient coverage of the 5' end of the message. Thus, the random hexamers offer the advantages of a simple protocol that yields a product that can be used for amplification in multiple PCRs. Note that controlled synthesis of a second cDNA strand is unnecessary. However, including the alkaline hydrolysis step after the cDNA synthesis improves the quality of the final product as determined by agarose gel electrophoresis.
3. Contamination: One of the most pernicious problems associated with the extreme sensitivity of PCR is the potential for false amplifications as a result of contamination of the reactions by minute amounts of DNA. The most common source of contamination is the products of previous PCRs, and the best solution to the problem is extreme caution when handling the PCR reagents. To check for contaminants, a negative control reaction without any DNA template should always be run in parallel with any PCR. In the case of cDNA amplifications, two excellent negative controls are the omission of reverse transcriptase in the cDNA synthesis step, and alkaline hydrolysis of the RNA before the beginning of the procedure. Neither of these reactions should yield a PCR product.

A further source of contamination in cDNA amplifications is caused by the presence of genomic DNA. The simplest way to overcome this problem is to choose PCR priming sites that are separated by large introns so that only the spliced RNA sequences will be amplified. If the DNA and cDNA amplifications cannot be distinguished by primer positioning, then extra care should be taken during the preparation of the RNA to avoid collecting DNA. Consider DNAse treatment of the RNA only as a last resort.

4. Optimal PCR buffers: At least two PCR buffer systems are in common use at this time. We have had most experience with the DMSO-containing buffer described above *(10)*, but the buffer recommended by the Cetus Corporation (2.5 m*M* $MgCl_2$, 200 μ*M* dNTP, 50 m*M* KCl, 200 μg/mL gelatin, and 10 m*M* Tris-HCl at pH. 8.4) *(14)* works at least as well under most circumstances. The "Cetus buffer" has the potential disadvantage that the concentration of some of the ingredients may need to be carefully optimized to ensure most efficient and specific amplification. However, the "Cetus buffer" has the advantage that it is more likely to be compatible with subsequent procedures used to analyze the PCR products. This is sometimes evident when collecting PCR products by ethanol precipitation, when material other than DNA is sometimes pelleted from the DMSO-containing buffer (i.e., salt and protein). Whatever final PCR protocol is chosen, it is important that a high level of specificity is achieved in the amplification. PCRs that contain multiple species when analyzed by agarose gel electrophoresis usually do not sequence well.
5. Separate vs simultaneous amplification and single-strand production: A key step in the analysis is the generation of a single strand by asymmetric priming in a PCR-like reaction. As described in the original report of the method, a single PCR is performed with different amounts of each primer *(7)*. Initially there is an exponential increase in the amount of the desired fragment, and then, as one primer is exhausted, the second primer continues to produce single strands. We separate the two reactions, doing one PCR to generate plenty of double-stranded material, and then taking aliquot of the product to initiate a second reaction that contains only one primer. This is more cumbersome, but in our hands makes for more reliable results, presumably because the amount of double-stranded material is relatively constant when the single-strand production process begins.

   The two-step procedure also has the advantage that the success of the initial PCR can be monitored and that the PCR can be used to seed

multiple SSPRs. There is no need to return to the cDNA-synthesis products in each case. Increasing the distance between the primer used to generate single-stranded DNA template and the DNA-sequencing primer can diminish the signal from the sequenced products; however, primers as far as 4 kb apart have functioned reliably.

6. Intermediate steps between PCR and SSPRs: When initiating the SSPRs, it is not necessary to purify the products of the first reaction by phenol and NENsorb® chromatography, as has been previously described *(15)*. Instead, the second reaction can be initiated by simply taking a small aliquot of the first PCR (≤1%) without any purification. If more than 1% is used, then the products of the second reaction might not sequence well. If the SSPRs cannot be made to work this way, *then* try the phenol/NENsorb®-affinity-column approach.
   a. Dilute the PCR with an equal vol of H_2O.
   b. Extract with an equal vol of phenol (saturated with TE buffer).
   c. Reextract the phenol phase with an equal vol of fresh H_2O.
   d. Remove all traces of phenol with ether.
   e. Remove all traces of ether.
   f. Passage the DNA through a NENsorb® column, eluting with 50% methanol.
   g. Lyophilize and resuspend in 50 µL of H_2O; use 1–2 µL for SSPR.
7. Agarose gel electrophoresis of SSPR products: In most cases, the analysis of SSPR products by agarose gel electrophoresis reveals a band at the position of the double-stranded fragment, and a faster-migrating band representing the single-stranded material (Fig. 2). At high agarose concentrations (>1%) or when the single strands have an unusual secondary structure, the single-strand band is sometimes at a position of higher mol wt. Not infrequently, multiple bands are seen, which may reflect the presence of many different secondary structures in the single strands or may be attributable to the internal priming of the double-stranded template during SSPR. All different types of SSPR product can sequence well, but there is a loose correlation between the complexity of the agarose gel morphology and the failure to sequence. In general, it is the "cleanliness" of the initial PCR that is more important than the agarose gel pattern of the SSPR product.
8. Buffers for SSPR: Either of the PCR buffer systems described above can be used for the SSPRs. However, when using the DMSO-containing buffer, it is particularly important that there is no carry-over of salt or protein into the pellet to be used for the DNA sequencing. Therefore, we have a preference for the use of the buffer recommended by the Cetus Corpo-

Fig. 2. PCR amplification of hypoxanthine phosphoribosyltransferase (HPRT) cDNA and production of single strands. A 920-b fragment containing the human peptide-coding region was amplified from cDNA as described in the text, using the specific oligonucleotide primers #365 (5'- CCG CCC AAA GGG AAC TGA TAG TC -3') and #863 (5'- CTT CCT CCT CCT GAG CAG TCA G -3'). Single strands were generated from the PCR products as described in the text. Lane M. W., mol wt markers; Lane 1, PCR product; Lane 2, SSPR product using oligimer #365; Lane 3, SSPR product using oligimer #863. The faster-moving bands represent the single-stranded fragments.

ration for the SSPRs, except when we find that a particular primer set functions much better in the DMSO-containing buffer. In that case, the DMSO-containing buffer is used in the SSPR, but great care is taken to avoid salt or protein coprecipitation.

9. Sequencing primers: The DNA-sequencing primers are routinely constructed as 18-mers. The use of end-labeled primers that are complementary to sequences between the PCR primers that were first used for amplification enables greater specificity, since nonspecific PCR contaminants will not be primed during the sequencing. However, the PCR primers usually can be used as the sequencing primers if the initial PCRs appear homogeneous when assayed by agarose gel electrophoresis. This is a great advantage, since it obviates the need for the construction of additional oligimers. To ensure that the PCR primers can be used for sequencing, we find it necessary to (a) use the minimum amount of primer in the initial PCR (as little as 5 pmol of each primer) and (b) perform the minimum number of PCR cycles that produce a visible band on an agarose gel from 10% of the reaction products (as few as 23 for rare cDNAs or unique human gene sequences). Thus, when trouble-

shooting a reaction in which the PCR primers will not give a good sequence, and when it is not desirable to synthesize a new oligomer, the amount of PCR primer and the number of cycles in the initial reactions should be titrated downward.

10. Sequencing buffer: The sequencing buffer that we prefer is the reverse-transcription buffer (*see* POL buffer, *above*) and not the usual mixture recommended by United States Biochemicals. The POL buffer has a lower ionic strength and, in our hands, gives a cleaner sequence.
11. Sequenase™ vs reverse transcriptase or *Taq:* Reverse transcriptase and *Taq* DNA polymerase have each been used for direct DNA sequencing. We have not had success with reverse transcriptase, although others report good results *(5)*. We have had no experience with *Taq,* but note that others report the superiority of that enzyme *(16)*. *Taq* sequencing is expensive, both because of the cost of the enzyme and because of the high concentrations of nucleotides that must be used. We have not encountered a region of DNA secondary structure that could not be resolved by T7 DNA polymerase sequencing at 50°C, and believe that the only advantage of *Taq* will be in the coupling of PCR to the sequencing by fully automated protocols *(15)*.
12. Sequencing reaction temperature: A most important feature of this protocol is the temperature of the sequencing reactions. In our hands the results from reactions at 50°C are spectacularly better than those from reactions at 37°C (*see* Fig. 3).
13. Automated DNA-sequencing: The direct DNA sequencing procedure can be automated by the use of fluorescent DNA sequencing primers (*see also* Chapter 4) and a commercially available fluorescent gel reader *(15,17,18)*. The manipulations for the automated DNA-sequence analysis are essentially the same as those for manual DNA sequencing. If necessary, the products of two SSPRs can be pooled before distribution of aliquots to be annealed to each of the custom-produced primers. Automated DNA sequencing of the PCR products routinely yields 275–350 bases of sequence, and it is likely that this can be extended by further "fine tuning" of the reaction conditions.

Acknowledgments

We thank Grant MacGregor for reviewing this manuscript. R. A. G. is a recipient of the Muscular Dystrophy Association's Robert G. Sampson Distinguished Research Fellowship, and C. T. C. is an investigator of the Howard Hughes Medical Institute. Supported by DHS grant #DK31428 and Welch Foundation grant # Q-533.

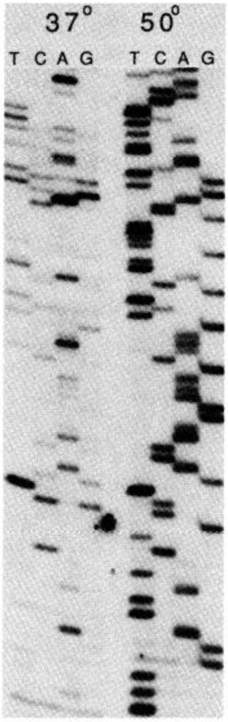


Fig. 3. Direct DNA sequencing with T7 DNA polymerase at 37 or 50°C. Two otherwise identical DNA sequencing reactions were performed at 37 or 50°C, according to the procedure described here.

References

1. Mullis, K. and Faloona, F. A. (1987) Specific synthesis of DNA *in vitro* via a polymerase catalyzed chain reaction. *Methods Enzymol.* **155,** 335–350.
2. Saiki, R. K., Scharf, F., Faloona, F., Mullis, K. B., Horn, G., Erlich, H. A., and Arnheim, N. (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230,** 1350–1354.
3. Wrischnik, L. A., Higuchi, R. G., Stoneking, M., Erlich, H. A., Arnheim, N., and Wilson, A. C. (1987) Length mutations in human mitochondrial DNA: Direct sequencing of enzymatically amplified DNA. *Nucleic Acids Res.* **15,** 529–542.
4. McMahon, G., Davis, E., and Wogan, G. N. (1988) Characterization of c-ki-ras oncogene alleles by direct sequencing of enzymatically amplified DNA from carcinogen-induced tumors. *Proc. Natl. Acad. Sci. USA* **84,** 4974–4978.
5. Wong, C., Dowling, C. E., Saiki, R. K., Higuchi, R. G., Erlich, H. A., and Kazazian, H. H. Jr. (1987) Characterization of beta-thalassemia mutations using direct genomic sequencing of amplified single copy DNA. *Nature* **30,** 384–386.

6. Yandell, D. W. (1989) Direct genomic sequencing of alleles at the retinoblastoma locus· Applications to carrier diagnosis and genetic counselling, in *Cancer Cells: Molecular Diagnostics of Human Cancer,* vol 7, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 223–227.
7. Gyllensten, U. B. and Erlich, H. (1988) Generation of single stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus *Proc Natl. Acad Sci. USA* **85,** 7652–7656.
8. Winship, P. R. (1989) An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. *Nucleic Acids Res.* **17,** 1266
9. Carothers, A. M , Urlaub, G , Mucha, J., Grunberger, D , and Chasin, L. A (1989) Point mutation analysis in a human gene: Rapid preparation of total RNA, PCR amplification of cDNA, and *Taq* sequencing by a novel method *Biotechniques* **7,** 494–499
10. Kogan, S. C., Doherty, M., and Gitschier, J. (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences Application to hemophilia A. *N. Engl. J. Med.* **317,** 985–990
11 Gibbs, R. A , Chamberlain, J. S , and Caskey C. T. (1989) Diagnosis of new mutation diseases using the polymerase chain reaction, in *The Polymerase Chain Reaction Principles and Applications* (Erlich, H , ed.), Stockton, New York, pp. 171–191
12. Kawasaki, E (1989) Detection of gene expression, in *The Polymerase Chain Reaction. Principles and Applications* (Erlich, H., ed.), Stockton, New York, pp. 89–97
13 Chirgwin, J. M., Przybyla, A. E., McDonald, R J , and Rutter, W J (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18,** 5294–5299.
14. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S J , Higuchi, R , Horn, G T., and Mullis, K. B. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487–491.
15. Gibbs, R. A., Nguyen, P. N., McBride, L J , Koepf, S. M , and Caskey, C T (1989) Identification of mutations leading to the Lesch-Nyhan syndrome by automated direct DNA sequencing of *in vitro* amplified cDNA *Proc Natl. Acad Sci USA* **89,** 1919–1923.
16 Innis, M A , Myambo, K. B , Gelfand, D. H., and Brow, M A (1988) DNA sequencing with *Thermus acquaticus* DNA polymerase and direct sequencing of polymerase chain reaction amplified DNA *Proc. Natl. Acad. Sci. USA* **85,** 9436–9440
17 McBride, L J., Koepf, S. M , Gibbs, R A , Nguyen, P N , Salser, W , Mayrand, P E , Hunkapiller, M. W., and Kronick, M. N. (1989) Automated DNA sequencing methods using polymerase chain reaction *Clin. Chem* , **35,** 2196–2201
18. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C , Connell, C. R., Heiner, C., Kent, S. B H., and Hood, L. E. (1986) Fluorescence detection in automated DNA sequence analysis *Nature* **321,** 674

CHAPTER 3

Direct Sequencing of PCR-Amplified DNA

Peter M. Green and Francesco Giannelli

1. Introduction

The polymerase chain reaction (*see* Chapter 1) allows the rapid isolation of specific DNA targets that may be used as sequencing templates either directly or after cloning into M13. The latter procedure allows single-strand sequencing, but is otherwise undesirable not only because it is slow, but also because a significant proportion of the amplified DNA molecules contain replication errors. These are expected to occur at a frequency of 1 in 10,000 bases incorporated *(1),* and will also be amplified during subsequent cycles. This means that at least three clones from independent amplification experiments must be sequenced in order to identify these replication errors and determine the final consensus sequence. Direct sequencing of the PCR product bypasses this problem since it produces an "average sequence" of all the copies of the target, and any miscopied molecule is bound to represent only a very small proportion of the total (unless one starts PCR with very few molecules of the target DNA).

The technique described here for the direct sequencing of PCR products is based on the "traditional" dideoxynucleotide (ddNTP) sequencing method developed by Sanger, Nicklen, and Coulson in 1977 *(2)*. The procedure uses a modified T7 DNA polymerase, Sequenase™ (USB), in place of the Klenow enzyme, and ddNTPs to terminate specifically DNA synthesis at either A, C, G, or T in such a way as to produce a population of molecules where every possible length is represented in sufficient amounts to be detected by autoradiography after fractionation on a denaturing polyacrylamide gel. This results in a "ladder" of bands across four tracks that are read

From: *Methods in Molecular Biology, Vol 9. Protocols in Human Molecular Genetics*
Edited by: C Mathew Copyright © 1991 The Humana Press Inc , Clifton, NJ

upward to give the sequence of a particular template DNA. When sequencing large numbers of products with the same primer, it is more useful to load all the "A" tracks adjacent, then the "Cs," and so on, and look for pattern changes that would indicate a mutation *(3)*.

However, direct sequencing of PCR products requires first of all the elimination of primers, dNTPs, PCR buffer, and *Taq* polymerase, since these would interfere with sequencing. This is done by binding the DNA to a glass bead suspension (GeneClean from Bio 101), washing it, and then eluting it in a small vol. The DNA binds to the "Glassmilk" suspension while other ingredients are washed away. The oligonucleotides appear to bind too tightly to be efficiently eluted.

2. Materials

1. GeneClean kit (Bio 101). This contains all the ingredients needed to purify the PCR products prior to sequencing, i.e., sodium iodide (saturated solution), "Glassmilk" suspension, and "NEW" wash buffer.
2. TE: 10 m*M* Tris-HCl, pH 8, 0.1 m*M* EDTA.
3. Sequenase™ (USB). This is the trade name for a modified T7 DNA polymerase. It should be noted, however, that the sequencing strategy of the Sequenase kit differs from that described here and is not very useful for sequencing PCR products.
4. dNTP/ddNTP mixes:

	Mix, µL			
	A°	C	G°	T°
Stock solutions				
0.5 m*M* dCTP	80	80	80	80
0.5 m*M* dGTP	80	80	80	80
0.5 m*M* dTTP	80	80	80	80
0.05 m*M* ddATP	0.8			
0.05 m*M* ddCTP		80		
0.05 m*M* ddGTP			80	
0.05 m*M* ddTTP				80
TE	260	180	180	180
Final vol =	500	500	500	500

5. 5× Sequenase™ buffer: 200 m*M* Tris-HCl, pH 7.5, 100 m*M* $MgCl_2$, 250 m*M* NaCl.
6. Dimethyl sulfoxide (DMSO): Freeze in 100-µL aliquots.
7. $\alpha[S^{35}]$dATP (600 Ci/mmol). Store in 4-µL aliquots at –70°C.
8. 100 m*M* Dithiothreitol (DTT): Freeze in 100-µL aliquots.

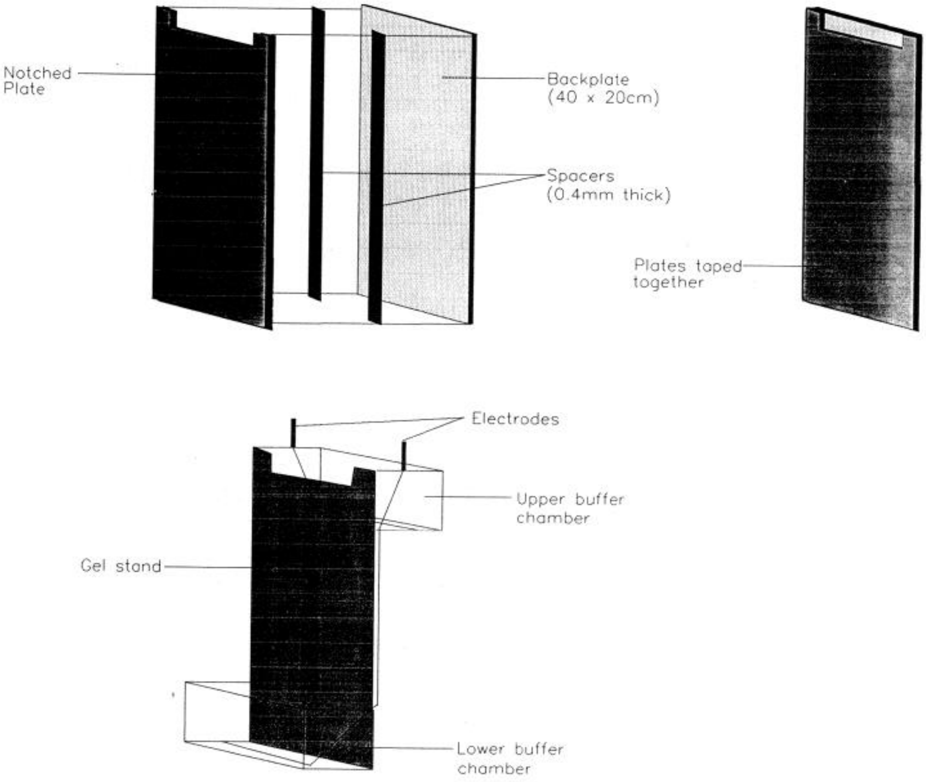


Fig. 1. Diagram of apparatus for polyacrylamide gels. The equipment should include a safety lid that completely covers the gel stand to protect against shocks, and an aluminum plate to clamp on front of the gel plates as a heatsink.

9. Oligonucleotide primers at 100 ng/µL: Either the primers used for PCR or primers internal to the product can be used.
10. Microtiter plates: U-shaped wells are best and they must be resistant to boiling (e.g., Nunc).
11. Chase solution: 0.25 m*M* dATP, 0.25 m*M* dCTP, 0.25 m*M* dGTP, 0.25 m*M* dTTP, 10% DMSO.
12. Gel loading dyes: 10 mg Bromophenol blue; 10 mg xylene cyanol; 30-mL deionized formamide.
13. Polyacrylamide gel electrophoresis. The basic apparatus for running sequencing gels is shown in Fig. 1. It can either be bought from various suppliers or constructed in a university or hospital workshop.
14. "Repelcote" (BDH) or similar siliconizing solution.

15. 40% Acrylamide stock solution: 76 g acrylamide, 4 g *N,N*-methylene-bisacrylamide; made up to 200 mL with distilled water. Deionize for 30 min by stirring with ~10 g "Amberlite" mixed bed resin (BDH). Filter the solution into darkened bottles and store at 4°C. Acrylamide is highly toxic and should be prepared in a fume hood with gloves, goggles, and a face mask when weighing out solid. It is possible to buy ready-to-pour gel mixes from several companies. This is safer and convenient but expensive!
16. 10% Ammonium persulfate (APS). Make up 10 mL at a time; keep at 4°C for up to 1 mo.
17. *N,N,N',N'*-Tetramethylethylenediamine (TEMED).
18. 10× TBE buffer: 0.89*M* Tris-base, 0.89*M* boric acid, 0.02*M* EDTA.
19. Gel fixing solution: 10% Acetic acid; 10% methanol. Make up 1L that can be reused a number of times.
20. X-ray film (e.g., Kodak XS-1) and cassettes.
21. Along with standard laboratory equipment, the following are useful: gel drier, salad spinner, and Hamilton repeater syringe (for dispensing 2 µL repeatedly with a yellow tip).

3. Method

3.1. Gene Cleaning of PCR Product

1. After removing the paraffin oil, add 2.5 vol of saturated sodium iodide (supplied with kit) to 1–2 µg of PCR product in a 0.5-mL Eppendorf tube.
2. Add 5 µL of the "Glassmilk" suspension, vortex, and leave for 5 min at room temperature.
3. Spin tubes for 15 s at full speed (~14,000 rpm) in a microcentrifuge tube.
4. Remove and discard the supernatant with a yellow tip.
5. Add 200 µL of "NEW" wash buffer (supplied with the kit and kept at –20°C).
6. Vortex and spin for 15 s. Repeat steps 4–6 twice.
7. After removing the supernatant, respin the pellet for 15 s and remove residual liquid (including any remaining paraffin) with a drawn-out Pasteur pipet.
8. Add 5 µL of TE to the pellet and resuspend with the automatic pipet. Incubate for 5 min at 55°C (either in a water bath or a PCR machine).
9. Spin for 30 s. Transfer the supernatant to a fresh 0.5-mL Eppendorf tube.
10. Repeat Steps 8 and 9. Combine the supernatants to give 10 µL of purified PCR product. If desired, run 1 µL on a gel to check recovery (should be 80–90% or higher).

3.2. Sequencing Reaction

1. In order to sequence eight templates simultaneously with the same primer, make up the following primer premix: 5 µL Primer (at 100 ng/µL), 18 µL 5x Sequenase buffer, 25 µL TE, and 6 µL DMSO (*see* Note 3).
2. Aliquot 6 µL of this primer premix to eight 0.5-mL Eppendorf tubes and add 1 µL of each purified PCR product ("template") per tube, mixing with a yellow tip each time. Put these tubes to one side for a few minutes.
3. Dispense 2 µL of each dNTP/ddNTP mix into four microtiter wells, for each template with the Hamilton repeater. Mark the plate by template number (1–8) and nucleotide mix (A°, C°, G°, and T°).
4. Make up the enzyme/label premix: 4 µL α[S^{35}]dATP (600 Ci/mmol), 8 µL 0 1*M* DTT, 19 µL TE, 3 µL DMSO, and 1 µL Sequenase (12.5 U).
5. Heat the primer/template mixes (from Step 2) to 95°C for 5 min (either in a boiling water bath or a PCR machine). Snap-cool on an ice/water bath.
6. Add 4 µL of the enzyme/label mix (from Step 4) to the side of the tubes containing the primer/template mixes. Flick-spin to mix.
7. Use the Hamilton repeater to dispense 2 µL from tube 1 to each of the four microtiter wells labeled 1-A, 1-C, 1-G, and 1-T. Repeat for template/primer mixes 2–8.
8. Use the salad spinner to spin down the drops in a microtiter plate. Alternatively, tap on the bench to knock the droplets down to the bottom.
9. Put tape around the edge of plate and float on 37°C water bath for 5 min.
10. Add 2 µL of chase solution to every well, spin to mix, and again incubate at 37°C for 5 min.
11. Add 2 µL of running dyes to each well, and spin to mix.
12. When the gel is ready for loading *(see below)*, heat to 95°C for 3 min by floating on a "simmering" water bath (fast boiling will flood the plate).
13. Snap-chill on ice/water. Samples are now ready for loading on the gel.

3.3. Polyacrylamide Gel Electrophoresis

1. The glass plates (Fig. 1) must be cleaned thoroughly in soap and water, and then dried with paper towels and ethanol.
2. Use "Repelcote" to siliconize the notched plate only. Clean with ethanol. Repeat if this is the first use of the plate.
3. Tape the plates together with the 0.4-mm spacers down each side, ensuring that the tape sticks firmly all round, especially at the bottom.
4. Make up the gel mix: 75 mL 40% Acrylamide, 25 mL 10x TBE, and 230 g urea; make up to 500 mL with distilled water.

5. Measure out 40 mL of gel mix (for a 40-cm × 20-cm × 0.4-mm gel) into a beaker. Add 140 µL of 10% APS, and then 90 µL of TEMED. Mix.
6. Draw up gel mix in a 50-mL syringe (without a needle). Holding the plates at a 45° angle, carefully squirt the acrylamide between the plates adjacent to one of the spacers. Take care not to get any air bubbles (sharp tapping on the glass can dislodge bubbles if it is done immediately). As the liquid nears the top of the plates, they should be gently lowered so that the top end rests on, for example, a 1.5-mL Eppendorf tube (i.e., it is left to set with the top end slightly higher than the bottom). Place the comb carefully in the top and leave to set for about 30 min. Any remaining acrylamide in the beaker should be left as an indication of when the gel has set. The higher the room temperature, the faster this will be.
7. Once set, the tape along the bottom of the gel should be removed and the gel placed in the apparatus with the backplate outermost. Clamp on the heatsink.
8. Fill the upper and lower chambers with 0.5× TBE running buffer.
9. Once the wells are covered in running buffer, remove the comb carefully. Flush out any nonpolymerized acrylamide and urea with a Pasteur pipet. Put on the gel cover, plug in, and preelectrophorese the gel for about 30 min at 30 W constant power.
10. Before loading the samples, make sure the wells are flushed out with a Pasteur pipet to remove any urea that may have leached out of the gel. Load the samples with a drawn-out capillary and mouthpiece, or a specially flattened tip that can fit between the plates (e.g., from Gilson). If the same primer was used for all PCR products, load all "A" reactions side by side, then the "C" reactions, followed by the "Gs" and "Ts".
11. Run the gel at 30 W constant power or whatever is required to get a gel temperature of about 55–60°C—this helps keep the products denatured. On a 6% gel, the bromophenol blue runs with the primers, whereas the xylene cyanol runs at about 70–80 bp from the primer. For a short run, let the bromophenol blue run just off the end of the gel. For a longer run, allow the xylene cyanol to run to the end of the gel (*see* Note 2).

3.4. Gel Processing and Autoradiography

1. Remove the gel from the stand, and remove all tape and spacers. Gently prize the plates apart taking care not to chip or crack them.
2. The gel should stick to the back plate (the nonsiliconized one). Put this in a large tray and carefully pour on the fixing solution. Leave for 30 min.

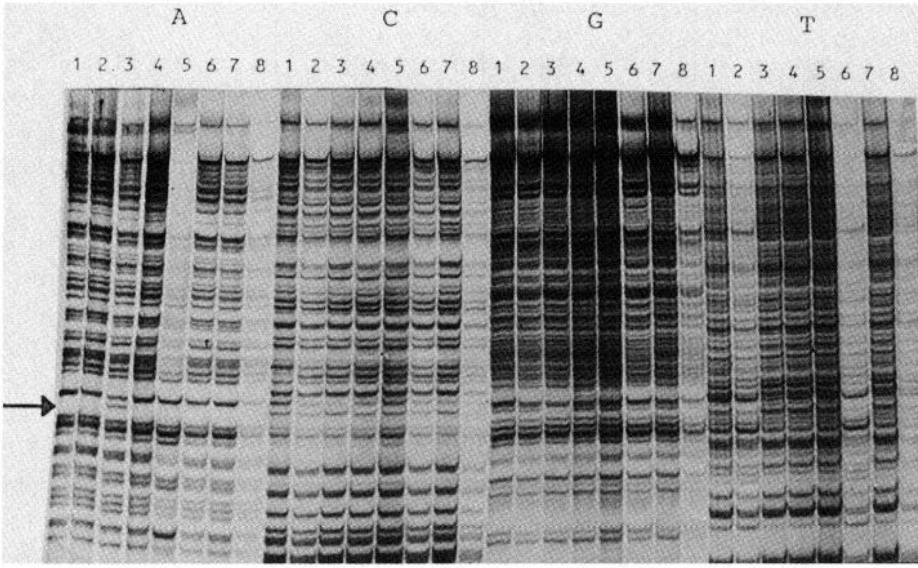


Fig. 2. Autoradiograph of sequencing gel. The same DNA segment of the factor IX gene from eight hemophilia B patients was sequenced, and samples were run in parallel. The arrow indicates an extra band seen in track A3. At the same level in the gel, a band is absent in track G3. Thus, this patient has a G→A transition at this position.

3. Gently lift the plate out of the fixing bath and place gel-up on paper towels. Cut two pieces of Whatman 3MM paper to approximately the size of the gel and lay on the gel.
4. Flip the paper–gel–plate "sandwich" over, so that the glass plate is uppermost. Lift the plate off the gel, which should stick firmly to the paper. Any wrinkles can be smoothed out with a wetted gloved finger.
5. Cover the gel with Saran Wrap™ and dry under vacuum in the gel drier at 80°C (*see* Note 4). Depending on the strength of the vacuum, this will take 20–60 min.
6. Remove the Saran Wrap™ from the dried gel. Place in an autoradiography cassette with "slow" X-ray film (e.g., Kodak XS-1, Fuji RX). Intensifying screens are of no benefit with ^{35}S. Expose at room temperature for 1–14 d.
7. Develop according to manufacturer's instructions. An example of part of a sequencing gel autoradiograph is shown in Fig. 2. A mutation is clearly visible even quite high up the gel.

4. Notes

1. The sequencing primer can be either one of the two used for PCR, or an internal one. The advantage of using an internal primer is that sequencing is usually successful even if the PCR product is not pure, e.g., unwanted bands appearing on the agarose gel. They may, of course, be necessary for longer products.
2. Often, the first 20–50 bp of sequence are unreadable. For this reason, it is usually best to run gels so that the xylene cyanol reaches the end of the gel. The sequence of the first 20–50 bp can be determined from the complementary strand using a primer that will extend in the opposite direction.
3. In place of DMSO, the detergents NP40 and Tween-20 at 0.5% final concentration have been used to reduce secondary structure and reannealing of the two strands *(4)*.
4. Gel driers are expensive. An alternative is to treat the back plate with "silane" by rubbing into the glass plate the following solution: 2.5 mL ethanol, 140 µL of 10% acetic acid, 7.5 µL of methacryloxypropyltrimethoxysilane (Sigma). The plate is then cleaned vigorously with ethanol. This procedure will allow the gel to bond tightly to the plate, which, at the end of the run, can be baked in the oven at 80°C. The gel dries to a thin film on the plate and is then autoradiographed. The plates are eventually cleaned by soaking in strong detergent overnight to remove the gel.

References

1. Saiki, R. K., Gelfund, D. H., Stoffel, S , Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487–491
2. Sanger, F., Nicklen, S , and Coulson, A. R (1977) DNA sequencing with chain-terminating inhibitors *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
3. Green, P. M , Bentley, D R , Mibashan, R S., Nilsson, I M , and Giannelli, F (1989) Molecular pathology of haemophilia B. *EMBO J* **8,** 1067–1072.
4 Bachmann, B , Luke, W , and Hunsmann, G. (1990) Improvement of PCR amplified DNA sequencing with the aid of detergents *Nucleic Acids Res* **18,** 1309

CHAPTER 4

Rapid DNA Sequence Analysis Using Fluorescent Labels

Richard K. Wilson

1. Introduction

Normal and disease-associated gene sequences may be rapidly and accurately characterized at the molecular level using the procedures described here. First, a modification of the polymerase chain reaction (PCR) technique *(1,2)* provides a simple method of template preparation starting from either genomic or cloned DNA samples. This modification, called asymmetric polymerase chain reaction (APCR), is diagrammed in Fig. 1. After a simple purification procedure, the resulting DNA is directly sequenced using an oligonucleotide primer labeled with a fluorescent reporter group. This preparation scheme eliminates the requirement of overnight culturing of bacteria or phage and provides the user with a rapid means of purifying sufficient template DNA for several sequencing reactions. The fluorescent DNA-sequencing procedure described here has been optimized to give the best results with the high-throughput APCR technique.

Recently, instrumentation that permits real-time detection of fluorescent-labeled DNA-sequencing reaction products has become available *(3,4)*. The advantages of this system include nonradioactive detection and elimination of manual autoradiograph interpretation. The system described here is manufactured by Applied Biosystems Inc. (Foster City, CA) This instrument is compatible with two types of fluorescent-labeling chemistries: (a) reporter group at the 5' terminus of the sequencing primer, and (b) reporter group at the 2' carbon of the dideoxynucleotide. Since the end-labeled fluorescent

From. *Methods in Molecular Biology, Vol. 9 Protocols in Human Molecular Genetics*
Edited by. C Mathew Copyright © 1991 The Humana Press Inc , Clifton, NJ

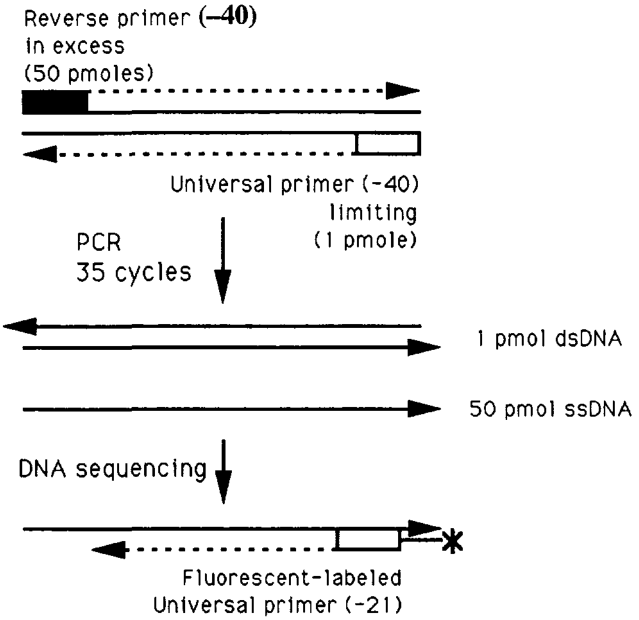


Fig. 1. APCR is used to rapidly prepare template DNA for nucleotide sequence analysis. APCR may be performed using single- or double-stranded DNA, bacteriophage plaques, or bacterial colonies. Here, the –40 universal and reverse primers are used to amplify an insert cloned in pUC 18 or 19. No purification of the resulting single-stranded template DNA is necessary.

chemistry currently gives the best results, it is the only method described here. An example of results from this automated system using the APCR and optimized sequencing methods is shown in Fig. 2. Other instruments that may be used for real-time, nonradioactive detection of DNA sequence reaction products have also been described *(5–7)*.

Prior to attempting these methods, the researcher should gain a thorough knowledge of the automated fluorescent DNA-sequencing system. The user's manual provided by the manufacturer is an important source of supplementary information.

2. Materials

1. A thermal cycler of some sort is required for the APCR procedure. These are available from several manufacturers, including Perkin-Elmer/Cetus (Emeryville, CA), Ericomp Inc. (San Diego, CA), and MJ Research

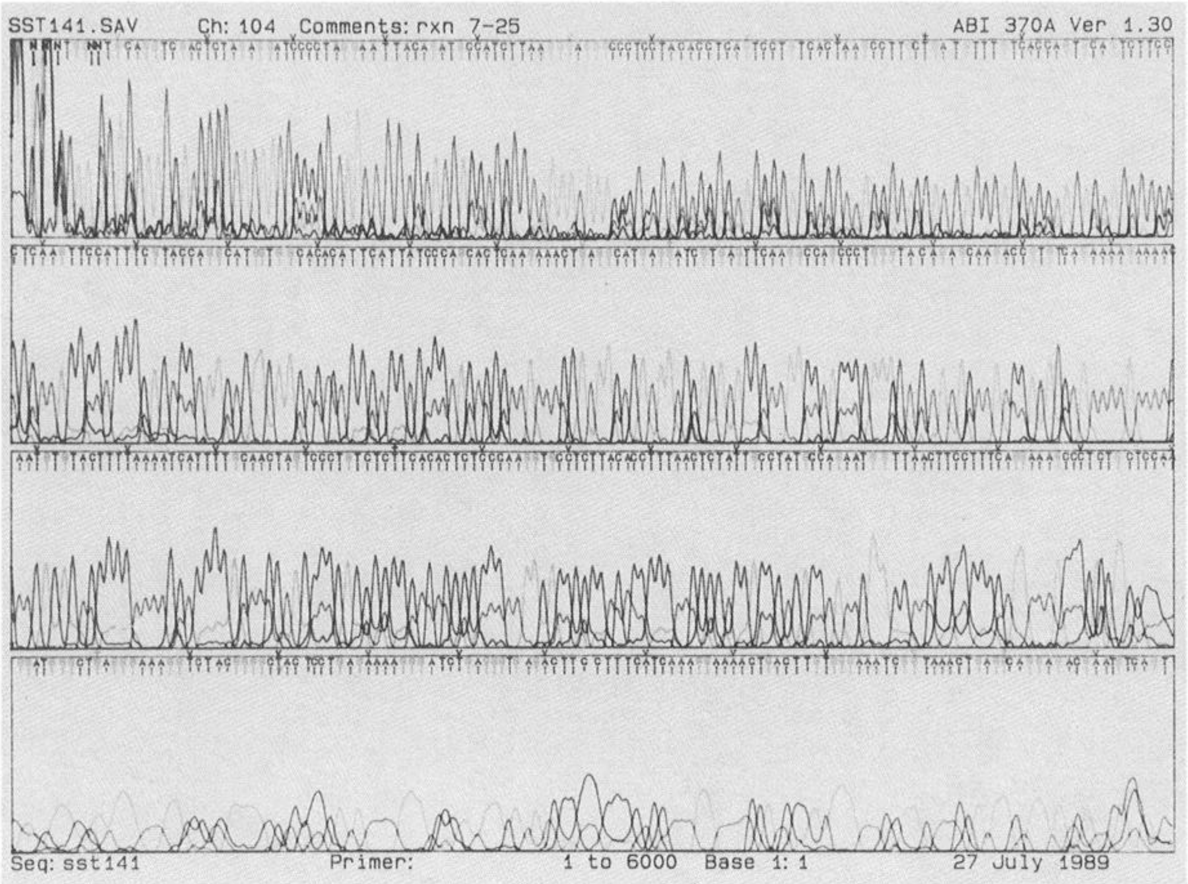


Fig. 2. Fluorescent DNA sequencing data obtained using direct APCR amplification of a recombinant M13 subclone from the mouse T-cell receptor a-chain locus. In this experiment, amplification was performed directly from a bacteriophage plaque using the –240 flanking primers. Here, the four-color chromatogram produced by the automated DNA sequencer is reproduced in black and white.

Inc. (Boston, MA). Simple homemade systems for thermal cycling have also been described *(8)*.

2. The automated fluorescent DNA-sequencing system described here is model 373A, manufactured by Applied Biosystems Inc. The current configuration includes an Apple® Macintosh® IIcx computer with an 80-MB hard disk drive.
3. *Thermus aquaticus (Taq)* DNA polymerase: This may be purchased from one of several enzyme suppliers; however, it is our experience that the enzyme supplied by Cetus gives the best results. For DNA-sequencing reactions, the enzyme of choice is the modified bacteriophage T7 DNA polymerase (US Biochemicals, Cleveland, OH) *(9)*.
4. Deoxynucleotides (dNTPs) and dideoxynucleotides (ddXTPs) are purchased from Pharmacia (Piscataway, NJ). For APCR, prepare a stock solution containing 1.25 m*M* of each dNTP.
5. For DNA-sequencing reactions, prepare a stock solution containing 8 m*M* of each dNTP. Prepare 50 μ*M* stocks of each ddXTP. All the nucleotide stock solutions should contain Tris-HCl, pH 7.6, at a final concentration of 10 m*M*, and should be stored at –20°C.
6. PCR buffer should be prepared as a 10× stock solution, according to the supplier's specifications.
7. Oligonucleotide primers for APCR should be 17–23 nucleotides in length. For optimal amplification of inserts cloned in M13 vector, the following primers should be employed: (–240 universal primer [UP]) 5' GGACGACGACCGTATCGG 3' and (–240 reverse primer [RP]) 5' GAAAAACCACCCTGGCGC 3' *(10)*. For optimal amplication of inserts cloned in pUC vectors, the following primers should be employed: (–40 UP) 5' GTTTTCCCAGTCACGAC 3' and (–40 RP) 5' GGATAACAATTTCACA 3' *(14)*. The primers should be kept as 12 μ*M* stocks at –20°C.
8. Fluorescent dye–primers are purchased from Applied Biosystems Inc. Since a double-stranded DNA is an intermediate of the APCR, DNA sequencing reactions may be performed from either end of the template DNA strand.
9. Stock solution for polyacrylamide gels: Prepare a 40% solution containing 38% acrylamide and 2% bisacrylamide. Store at 4°C.
10. 10× TEB electrophoresis running buffer: 1.33 *M* Tris-HCl, pH 7.6, 0.45 *M* boric acid, 25 m*M* EDTA. Store at 4°C.
11. 15% Ammonium peroxysulfate. Store at 4°C.
12. 5× Sequencing buffer: 50 m*M* Tris-HCl, pH 7.6, 15 m*M* $MnCl_2$, 300 m*M* NaCl, 5 m*M* dithiothreitol. Prepare fresh (*see* Note 4).
13. Loading solution: 0.25 *M* EDTA in deionized formamide.

14. 5 *M* Ammonium acetate, pH 7.4.
15. Enzyme dilution buffer (prepare according to supplier's specifications).
16. TE buffer: 10 m*M* Tris-HCl, pH 8.0, 0.1 m*M* EDTA.

3. Methods

3.1. APCR Template Preparation (see *Note 2*)

The APCR method allows rapid preparation of template for DNA sequencing reactions starting from bacteriophage plaques or bacterial colonies. For single copy genes, this method also should be applicable to genomic DNA.

1. If the DNA to be analyzed is subcloned in a plasmid or phage vector, a small amount of a plaque or colony containing the recombinant subclone is cored and lifted from the culture plate using a sterile Pasteur pipet. The agar plug then is dropped into 50 µL of sterile water and vortexed to release the bacteria and phage particles. The 50-µL samples should be kept at 4°C.
2. Prepare a master mix for APCR as follows:

sterile, distilled water	58 × *n* µL
10× PCR buffer	10 × *n* µL
1.25 m*M* dNTPs	16 × *n* µL
primer 1 (12 µ*M*)	5 × *n* µL
primer 2 (0 24 µM)	5 × *n* µL
Taq DNA polymerase (5 U/µL)	0.2 × *n* µL

where *n* equals the number of subclones or samples to be amplified. *Important:* the primer in excess (primer 1) corresponds to the template strand that will be analyzed in the subsequent DNA-sequencing reactions. For example, if the fluorescent-labeled –21 universal primer is to be used for sequence analysis, the limiting primer in APCR from M13 subclones should be –240 UP.

3. Aliquot the master mix to 0.6-mL microcentrifuge tubes in 95-µL amounts. To each tube, add 5 µL of the phage or colony stock from step 1. If DNA samples other than plaques or colonies are to be used, the appropriate amount (i.e., 10 ng of purified template or 1 µg of genomic DNA) should be added. Mix the solutions gently and overlay each solution with 50 µL of light mineral oil. Cap the tubes tightly and place in the automated thermal cycler.
4. Perform APCR for 35 cycles. Typically, denaturation is at 94°C for 40 s, followed by annealing at 55°C for 40 s. Primer extension by the *Taq* DNA

polymerase then proceeds at 72°C for 90 s. These conditions are sufficient to amplify DNA inserts of up to 4000 bp in length.

5. At the conclusion of APCR, remove the sample tubes from the thermal cycler and transfer the aqueous (lower) phase from each reaction to a clean 1.5-mL microcentrifuge tube. To each tube, add 10 µL of 3*M* sodium acetate, pH 5.2, and 200 µL of isopropanol. Mix briefly and let stand for 30 min at room temperature.
6. Pellet the DNA by centrifugation at 13,000*g* for 15 min at room temperature. Wash the DNA pellet with 400 µL of 70% ethanol, and dry briefly under vacuum. The dried DNA pellets should then be dissolved in 40 µL of TE buffer and stored at 4°C.

3.2. Fluorescent DNA-Sequencing Reactions

1. A 6% polyacrylamide gel containing 7*M* urea should be prepared using the glass plates, spacers, and well-forming combs supplied by Applied Biosystems. Detailed instructions for preparing, setting up, and prerunning the gel are provided in the user's manual.
2. DNA-sequencing reactions should be performed in 0.6-mL microcentrifuge tubes or, more conveniently, in 96-well V- or U-bottom microtiter plates (*see* Note 5). Care should be taken to keep the reactions away from fluorescent lighting. Set up four annealing reactions for each template DNA as follows.

	A	C	G	T
5× sequencing buffer	1 µL	1 µL	2 µL	2 µL
template DNA	3 µL	3 µL	6 µL	6 µL
dye–primer	1 µL	1 µL	2 µL	2 µL

Incubate the reactions at 55°C for 3–5 min, then cool slowly to room temperature over 15–30 min.

3. While the annealing reactions are incubating, label four 0.6-mL microcentrifuge tubes "A", "C", "G", and "T." To each tube, add equal vol of the 8 m*M* dNTP stock and the appropriate 50 µ*M* ddXTP. Be sure to prepare sufficient mix for all the reactions.
4. To each annealing reaction, add the following:

	A	C	G	T
8 m*M* dNTPs + 50 µ*M* ddXTP	2 µL	2 µL	4 µL	4 µL
mT7 DNA pol (1.5 U/µL)	1.5 µL	1.5 µL	3 µL	3 µL

Incubate at 37°C for 5 min.

5. Stop the reactions by adding EDTA as follows:

	A	C	G	T
0.025*M* EDTA	1 µL	1 µL	2 µL	2 µL

6. Combine the four reactions for each DNA template into one 1.5-mL microcentrifuge tube. Add 6 µL of 5*M* ammonium acetate, pH 7.4, and 120 µL of cold 95% ethanol. Alternatively, the reactions may be stopped by sequential transfer to the ammonium acetate/ethanol mixture, thereby eliminating step 5. Precipitate the fluorescent-labeled DNA at –70°C for 15 min.
7. Pellet the DNA by centrifugation at 13,000*g* for 15 min at 4°C. Wash with 400 µL of 70% ethanol and dry for a few minutes under vacuum. The dried samples may be stored at –20°C for several days.
8. Immediately before loading samples on the automated DNA sequencer, completely and carefully dissolve the DNA pellets in 5 µL of formamide–EDTA loading solution. Heat at 100°C for 3–5 min and then place on ice. Load each sample into single wells on the sequencing gel. Start the DNA sequencer and conduct the automated run as directed in the user's manual.

4. Notes

1. Currently, the Applied Biosystems automated fluorescent DNA sequencer requires approx 12 h to collect data for about 500 bp/sample. The above protocol as written has been optimized for this system.
2. For most template DNAs, the conditions described above for APCR amplification will be applicable. However, it has been our observation that occasional modifications are required. The simplest of these include increasing the amount of *Taq* DNA polymerase in the APCR mix, titrating the Mg^{2+} concentration in the PCR buffer, increasingor decreasing the annealing temperature, and increasing the denaturation (94°C) and extension (72°C) times.
3. As has been discussed elsewhere *(10)*, the high-throughput APCR template-preparation scheme with recombinant M13 clones requires some distance between the end of the single-stranded APCR product and the annealing site of the fluorescent-labeled sequencing primer. This is probably because of the presence of incomplete product strands in the APCR amplification. If templates other than recombinant M13 or pUC subclones are employed, this should be taken into account prior to the design of amplification and sequencing primers. Alternatively, the APCR product may be purified by biotin–streptavidin *(11)* or HPLC methods.
4. As described previously by Tabor and Richardson *(9)*, the replacement of Mg^{2+} with Mn^{2+} in the sequencing buffer improves the processivity of the T7 DNA polymerase, resulting in more accurate base assignment on

the automated fluorescent sequencer. It is important, however, that the Mn^{2+}-containing buffer be made fresh prior to each sequencing experiment. It is convenient to keep all components of the Sequencing buffer as 10× stocks.

5. The current capacity of the Applied Biosystems Inc. DNA sequencer is 24 samples. In order to simplify the task of performing the fluorescent DNA-sequencing reactions, the use of polystyrene or vinyl microtiter plates and an eight-channel micropipetor is highly recommended. If 24 sequencing reactions are to be performed simultaneously, it is suggested that the microtiter plate be placed on ice during the later pipeting steps. Alternatively, a robotic workstation may be used to automate the fluorescent DNA-sequencing reactions *(12,13)*.
6. As previously mentioned, the user's manual that accompanies the automated fluorescent DNA sequencer should be studied extensively during the setup, electrophoresis, and data analysis tasks. Be sure to carefully back up all your DNA sequence data files via network connections to another computer or on floppy disks before beginning the next day's experiments.

Acknowledgment

The author wishes to thank C. Chen for critical comments and technical support.

References

1. Saiki, R. K., Scharf, S., Faloona, F , Mullis, K. B., Horn, G. T , Erlich, H. A , and Kazazian, H (1985) Enzymatic amplification of β-globin sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230,** 1350–1354.
2. Gyllensten, U. B. and Erlich, H. A (1989) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQα locus. *Proc Natl. Acad. Sci. USA* **85,** 7652–7656.
3. Smith, L. M , Sanders, J. Z., Kaiser, R J., Hughes, P., Dodd, C., Connell, C., Heiner, C , Kent, S B. H., and Hood, L. E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* **321,** 674–679.
4. Connell, C. R , Fung, S., Heiner, C , Bridgham, J , Chakerian, V , Heron, E , Jones, B , Menchen, S , Mordan, W., Raff, M., Smith, L., Springer, J , Woo, S , and Hunkapillar, M. (1987) Automated DNA sequence analysis. *BioTechniques* **5,** 342–348
5. Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J , Cocuzza, A. J., Jensen, M A., and Baumeister, K. (1987) A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238,** 336.
6. Ansorge, W., Sproat, B. S , Stegemann, J., Schwager, C , and Zenke, M (1987) Automated DNA sequencing: Ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.* **15,** 4593–4602

7 Brumbaugh, J. A., Middendorf, L. R., Grone, D. L , and Ruth, J. L (1988) Continuous, on-line DNA sequencing using oligodeoxynucleotide primer with multiple fluorophores. *Proc. Natl. Acad. Sci. USA* **85**, 5610–5614.

8. Cao, T M (1989) A simple and inexpensive system to amplify DNA by PCR. *BioTechniques* **7**, 566, 567.

9 Tabor, S. and Richardson, C. C (1989) Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage T7 DNA polymerase and *Escherichia coli* DNA polymerase I. *Proc Natl. Acad Sci. USA* **86**, 4076–4080

10. Wilson, R K., Chen, C , and Hood, L. (1990) Optimization of asymmetric polymerase chain reaction for rapid fluorescent DNA sequencing. *BioTechniques* **8**, 184–189

11. Mitchell, L G and Merril, C R. (1989) Affinity generation of single-stranded DNA for dideoxy sequencing following the polymerase chain reaction *Anal Biochem.* **178**, 239–242

12 Wilson, R K , Yuen, A S , Clark, S. M., Spence, C , Arakelian, P and Hood, L (1988) Automation of dideoxynucleotide DNA sequencing reactions using a robotic workstation *BioTechniques* **6**, 776–787

13. Wilson, R K , Chen, C , Avdalovic, N , Burns, J , and Hood, L (1990) Development of an automated procedure for fluorescent DNA sequencing *Genomics* **6**, 626–634

14 Du, Z , Hood, L , and Wilson, R. K (1991) Automated fluorescent DNA sequence analysis of asymmetric PCR products. *Methods in Enzymology,* in preparation

CHAPTER 5

Detection of Mutations in DNA and RNA by Chemical Cleavage

R. G. H. Cotton

1. Introduction

This technique was developed to screen for point mutations, but deletions and insertions too small to be recognized by gel electrophoretic techniques are also detected. Whereas earlier techniques are able to detect point mutations, not all mutations were detected and/or the technique was not convenient or direct (reviewed in *1*). The chemical cleavage of mismatch method (CCM) rapidly and reliably detects all classes of point mutations *(2)*. Reference DNA probe is mixed with excess test DNA or RNA; the mixture is melted, and then cooled to allow reannealing and, thus, heteroduplex formation with mismatched or unmatched base pairs at the position of the mutation. Probe is modified at mismatched C and T bases by reaction with hydroxylamine and osmium tetroxide, respectively, and subsequently cleaved by piperidine treatment. Fragments are sized on gels (of the type needed for sequencing) to locate the point of cleavage and, hence, the mutation. In the case of point mutations, mismatched G and A bases will not be directly detected, but they are transposed to mismatched C and T bases, respectively, by use of probe of opposite sense for detection. However, matched bases adjacent or close to mismatched or unmatched bases become reactive by transmission of the distortion *(2,3)*, and can signal the presence of the mutation and hence allow indirect detection. This allows detection of insertions *(3)*. Unmatched C and T bases are also reactive, allowing detection of deletions. It should be emphasized that this is a screening method developed to avoid the need for sequencing kilobases of DNA to detect a single mutation. Once

From. *Methods in Molecular Biology, Vol. 9. Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc , Clifton, NJ

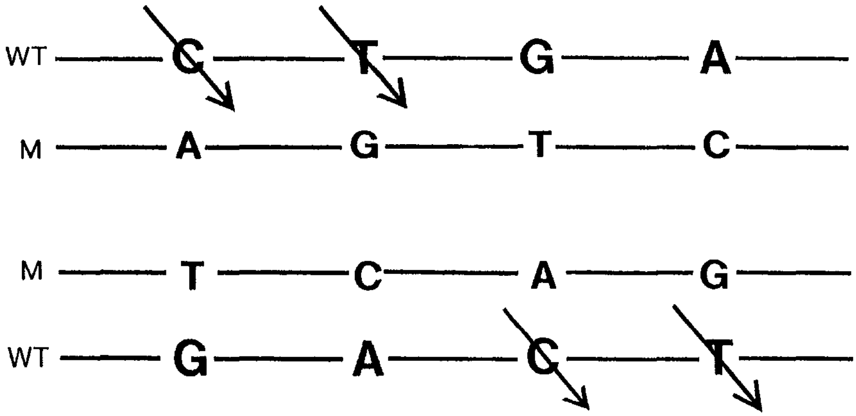


Fig. 1. All mutations can be detected by use of probes of both senses in making the heteroduplexes (top and bottom strands).

the site of the mutation is detected, only a small portion of the mutant gene needs to be sequenced. For subsequent detection in individuals, families, or populations, one of a series of simple mutation detection methods, such as oligonucleotide hybridization, can be used (*see* Chapters 7 and 24).

1.1. Strategy

Because all classes of C and T mismatches (C.C, C.T, C.A, T.T, T.G, and T.C) are cleaved *(2)*, complete screening of a double-stranded target for point mutations can be achieved using probes of both senses (Figs. 1 and 2a). Deletions will be detected by cleavage of unmatched C and T bases (Fig. 2c) or indirectly because of reactive bases nearby (Fig. 2d). Insertions will be detected indirectly by increased reactivity of nearby matched C and T bases in the probe next to or near the loop of the unlabeled DNA or RNA in the heteroduplex (Fig. 2f). In single-stranded targets, such as messenger RNA, increased reactivity of matched C and T bases near or next to the mismatched (Fig. 2b) or unmatched (Figs. 2d and 2f) bases become more important for complete screening. However, to be certain of detecting all classes of mutations directly, cDNA needs to be made for heteroduplex formation with probes of both senses.

The technique has two modes of use with either (a) uniformly labeled probe or (b) end-labeled probe. Either mode can be used when the variation expected is minimal, such as one mutation in the region covered by a probe. However, if multiple differences are expected, e.g., 1 base in 10 is likely to vary, an end-labeled probe will generate a single and unique band for each

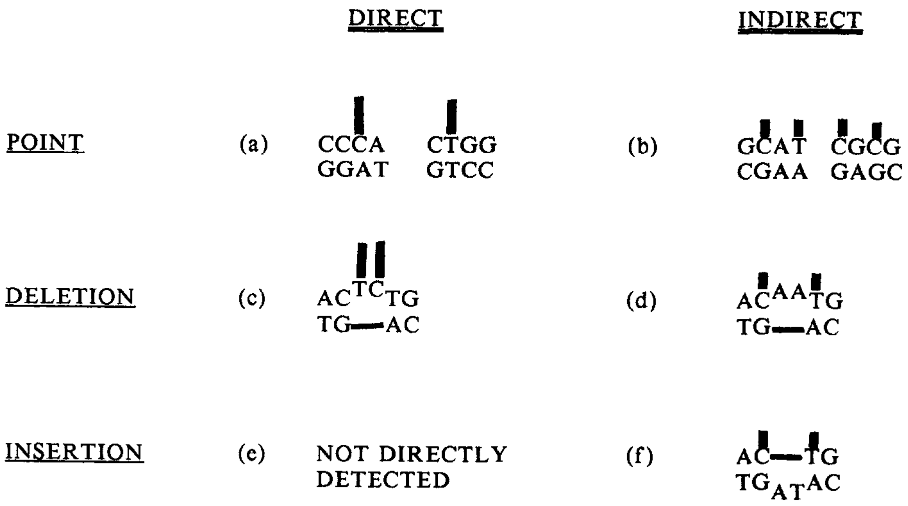


Fig. 2. Modes of detection of mutations. Heteroduplexes are shown with the probe strand in the upper position and test DNA or RNA in the lower position. Vertical lines indicate relative reactivity of T and C bases, with osmium tetroxide and hydroxylamine. Reactivity of matched nearby bases is less than unmatched or mismatched bases.

reactive C or T in the probe, thus generating a pattern of difference between the pieces of nucleic acid. When using end-labeled probe, the low reactivity of all matched C and T bases with hydroxylamine and osmium tetroxide, respectively, offers a background C and T track to help locate the mismatch.

The method has been applied to DNA amplified either via plasmids *(4)* or the polymerase chain reaction (PCR) technique *(5)*, and directly to RNA and viral RNA isolated from mammalian cells, but not as yet directly to unamplified genomic DNA. However, it should be noted that reactivity of T.U and T.G mismatches in DNA/RNA heteroduplexes has not yet been examined (reviewed in *1*), but they are expected to be reactive.

2. Materials

1. 2 × DNA/DNA annealing buffer: 1.2*M* NaCl, 12 m*M* Tris-HCl, pH 7.5, and 14 m*M* $MgCl_2$. Store at room temperature.
2. 5*M* Hydroxylamine solution: 1.39 g solid hydroxylamine hydrochloride or hydroxylammonium chloride is dissolved in 1.6 mL of warmed distilled water in a glass test tube; 1.75 mL diethylamine is added dropwise to bring the pH to 6. Store at 4°C for up to 1 wk.

3. 4% Osmium tetroxide solution: (N.B.: Osmium tetroxide irritates mucous membranes and should be used in a fume hood.) Break a 0.5-g ampule (Johnson Matthey, Materials Technology, Orchard Road, Royston, Herts 5685HE, England) and place in 12.5 mL distilled water in a glass bottle with a lid and a good seal. Stand for 2–3 d at 4°C to dissolve, and store. A fresh 1 in 5 dilution is made on the day of use.
4. 10× Osmium tetroxide buffer: 100 m*M* Tris-HCl, pH 7.7, 10 m*M* EDTA, 15% pyridine. Store at –20°C.
5. HOT stop buffer: 0.3*M* Na acetate, pH 5.2, 0.1 m*M* EDTA, 25 µg/mL tRNA. Store at –20°C.
6. DNA/RNA annealing buffer: 80 Parts filtered deionized formamide is added to 20 parts of the following, immediately before use: 200 m*M* PIPES, pH 6.5, 5 m*M* EDTA, and 2*M* NaCl. Filter for storage at room temperature.
7. Piperidine: 1*M* piperidine diluted immediately before use, from 10*M* stock. (N.B.: Piperidine is toxic and should be used in a fume hood.) Pyrrolidone is a noncontrolled alternative *(6)*.
8. Tris EDTA (TE): 20 m*M* Tris-HCl, pH 7.4, 1m*M* EDTA.

3. Methods

3.1. DNA and RNA Preparation

1. Prepare unlabeled DNA (about 10 µg) by standard methods, such as plasmid amplification *(4)* or PCR amplification *(5)*, and cut with appropriate enzymes.
2. Prepare unlabeled RNA (about 5 µg total cellular) by standard methods *(4)*. When RNA is abundant, e.g., collagen mRNA and viral RNA, total cellular RNA can be used as RNA source.
3. Prepare ^{32}P-labeled probe DNA by:
   a. Appropriate restriction enzyme digest of about 1 µg, so that 5' overhangs can be filled with Klenow fragment *(4)*. Prepare probes of both senses for the same region by filling 5' overhangs at both ends. Prepare single sense probes by making one cut with an enzyme not producing a 5' overhang that can be labeled, or if it does, cut off the label with an appropriate second enzyme. This provides single end-labeled probe.
   b. Filling in an M13 clone using the universal primer and radiolabeled base, and subsequently cutting out the insert *(2)*.
   c. Kinase labeling a PCR amplified fragment *(4)*.
   d. Adding radiolabeled bases during the PCR reaction *(7)*.

3.2. Heteroduplex Formation

3.2.1. DNA/DNA

1. Take labeled control and unlabeled test DNA (0.1–1 μg) in 20–100 μL TE and mix, so that there is at least a 12× excess of unlabeled DNA. To this, add an equal vol of 2× annealing buffer.
2. Place the mixture in a boiling water bath for 5 min and then transfer to 42°C for 60 min.
3. Precipitate the heteroduplex once with ethanol, wash once with 70% ethanol, and dry. Resuspend the DNA in distilled water, so that 1 μL contains 1000 dpm.
4. Prepare a homoduplex labeled control plus unlabeled control identically for the control reaction.

3.2.2. DNA/RNA

1. Mix DNA probe and test RNA (0.1–1 μg) so that there is at least a 12× excess of RNA, precipitate the mixture, wash with 70% ethanol, and dry the pellet.
2. Add 40 μL of DNA/RNA annealing buffer. Incubate at 90°C for 5 min, then at 55°C for 2 h, and then ethanol precipitate. Wash the pellet with 70% ethanol and dry. Resuspend the pellet in distilled water so that 1 μL contains 1000 dpm.

3.3. Modification Reactions

Osmium tetroxide and hydroxylamine tubes should be handled in a fume hood until ethanol precipitates.

1. Distribute 6000 dpm in 6 μL to each tube. Total DNA or DNA/RNA in each reaction should be 0.2–2.6 μg.
2. Hydroxylamine reaction (two tubes). Add 20 μL of hydroxylamine solution, mix, and incubate at 37°C for 10 and 30 min (partial cleavage), and 60 min for complete cleavage.
3. Osmium tetroxide reaction (two tubes). Add 2.5 μL of 10× osmium tetroxide buffer and 15 μL of diluted osmium tetroxide solution. Mix (do not centrifuge, as a yellow precipitate appears), and incubate at 37°C for 1 and 5 min for partial cleavage, and 20 and 60 min for complete cleavage.
4. Stopping reactions. Stop reactions with 200 μL of HOT stop buffer and 750 μL of ethanol and precipitate the DNA. Wash the pellet with 70% ethanol and dry.

3.4. *Piperidine Cleavage* (8)

Add 50 µL of 1*M* piperidine to each tube. Vortex for 10 s, heat at 90°C for 30 min, and cool on ice. Add 50 µL of 0.6*M* acetate buffer, pH 5.2 and 300 µL of ethanol to precipitate the DNA. Wash with 70% ethanol and dry.

Take up the samples with formamide dyes and analyze by electrophoresis on sequencing gels (*see* Chapter 3 and ref. *4*).

4. Notes

1. Probe length is limited only by the analytical technique and the fidelity of the heteroduplex formation. Probes up to 1.7 kb have been analyzed on sequencing gels.
2. Probe overlap. When screening kilobase lengths, probes need to be overlapped by 20–30 bases to avoid missing mismatches at the overlap. This is because of considerable breathing of the duplex at each end *(3)*.
3. Time of incubation. Pilot work needs to be done on the time needed for analysis of particular quantities of unlabeled DNA/DNA probe. If most of the probe remains uncleaved, there has been too little reaction, and if it is all cleaved, there may be too much reaction. This is because matched bases are modified and cleaved at a rate of about 1/100 of that of mismatched bases, allowing nonspecific probe destruction if the incubation time is too long. A time-course is necessary between these limits to be sure of a complete assessment (*see* Fig. 3).
4. Heteroduplex formation. Sometimes when changing operators or laboratories, it has been found that heteroduplexes have not been formed. This has been thought to be caused by the boiling water bath coming off the boil before melting, or use of too large a vol of solution for heteroduplex formation (>1 mL). Formation of a heteroduplex with a known mismatch should be included as a control that will be cleaved if in fact that heteroduplex has been formed.
5. Osmium tetroxide. Potency of this reagent between two laboratories has varied. It is not clear whether this is owing to different reagent batches or aging of an initial solution. A range of concentrations should be tried when setting up the method. Solutions should be freshly made each three months before the solution takes on a green hue, although it is still active at this time. Eppendorf tubes darken as a result of the reaction. Osmium tetroxide from one manufacturer has been ineffective, since it either does not cleave the T mismatches or the background is too high. Purity of the chemical that was successful from two different manufacturers was 99.8 and 99.9% (Aldrich and BDH, respectively).

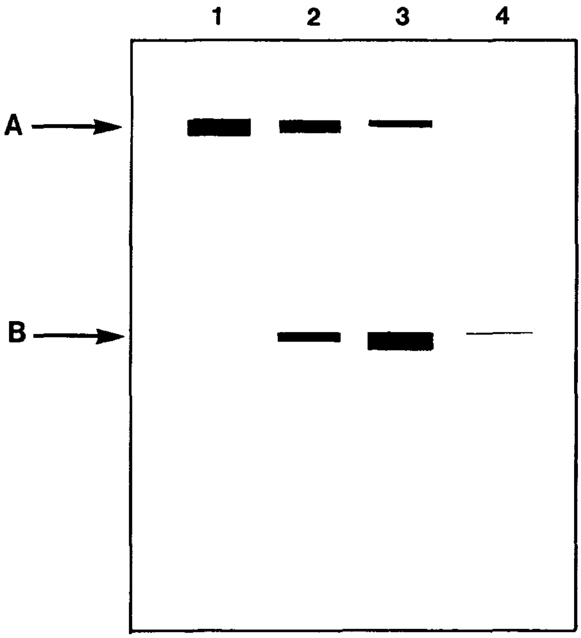


Fig. 3. Schematic representation of detection of a mutation in probe A hybridized to mutant DNA. Cleavage of the end-labeled probe by one of the HOT chemicals plus piperidine at a mismatch results in the fragment of reduced size (B) in the sequencing gel. Lanes 1–4 represent increasing chemical reaction time. The diagram illustrates that if too little incubation (lane 1) or too much incubation (lane 4) is given, band B (which signals the mutation) may not be seen.

In one case, the reagent was inactivated when the stock solution was prepared in a plastic bottle.

6. False positives and negatives. There have been no false positives identified so far. Some T.G mismatches have been seen whose reactivity is low or insignificant. However, it is fortuitous that these mutations, usually in a GC-rich region, will not be missed, since in two cases, the mutation was detected indirectly by the reactivity of nearby matched C bases. These had been made reactive by the unstable A.C mismatch (reviewed in *1*) using the probe of opposite sense (M. Anderson et al., in preparation).
7. Restriction enzyme sites have been added to PCR primers to facilitate probe labeling *(9)*.
8. Use with PCR and sequencing. The CCM method has been found to be especially useful with PCR-amplified products. This is because (a) in the

case of a heterozygote at a particular locus, one allele has been found to be more frequent in the product than the other, making interpretation of mutant sequence difficult after direct sequencing and (b) M13 cloning of PCR products from heterozygotes has, on occasion, shown many PCR errors, and there is a need to distinguish these from the actual mutation. Application of the CCM method samples the average of the product and indicates where to expect the mutations and which base is changed, i.e., distinguishes PCR artifacts from the real mutation.

9. Detection of homozygosity or heterozygosity. It is logical that cleavage kinetics should be different when mismatches are present in one or both alleles. Thus, in the case of homozygotes, the probe band should disappear to nothing at a uniform rate, but in the case of a heterozygous situation, the probe band should rapidly diminish to half intensity and then diminish more slowly and in parallel with the band signaling the mismatch. This situation is observed in the heterozygous mutation OI31 in Fig 4. However, a test has recently been developed *(10)* for testing allele status. This involves making the probe from the test DNA and hybridizing it with unlabeled test DNA. If the wild-type allele is present in the test sample, i.e., it is a heterozygote, cleavage will occur, but not if it is homozygous for the mutation in question.
10. Practical example. Several points are illustrated in Fig. 4. The major band at 320 bp represents direct detection of a mismatched C found in the heteroduplex as a result of a G→C mutation. It is notable that this mutation is heterozygous, since osteogenesis imperfecta is dominant. As noted above, the kinetics of band change is typical of a heterozygote. The minor band at 246 bp represents indirect detection of a polymorphism by reaction of a nearby matched C. Thus, such detection is easily possible when the mutation is only on one allele.
11. Criteria for CCM screening. Before a stretch of DNA can be deemed to be negative for the presence of mutations, it must appear identical with an identically prepared control homoduplex using the following criteria: (a) a range of reaction times are performed as per Fig. 3, (b) highest and lowest bands are visualized on long and short runs, respectively, and (c) heteroduplexes containing probes of both senses are treated with hydroxylamine and osmium tetroxide.
12. Detection of mutations in the heterozygous state. Detection of C mismatches from mutations present in the heterozygous state has not presented problems when probing with uniformly labeled probes of both senses in the same tube. However, recent data indicate that, as a result of the lower signal to noise ratio with osmium tetroxide and T mismatches

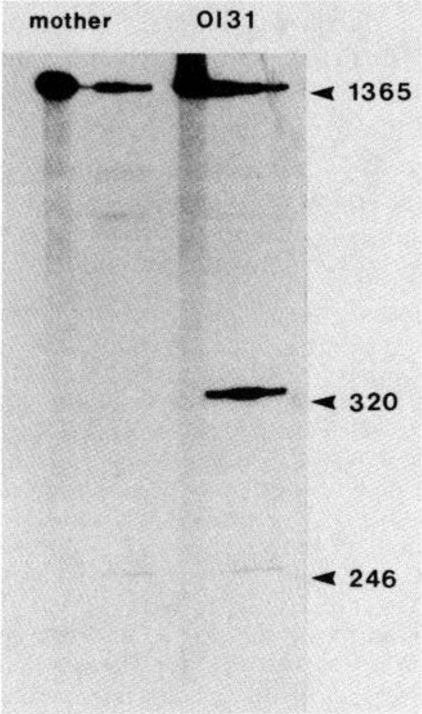


Fig. 4. Analysis of patient OI31 *(14)*. Fibroblast RNA was annealed to a 1365-bp end-labeled NcoI-EcoRI fragment from a human collagen α-l(I) cDNA clone Hf404. Treatment with hydroxylamine was as described in Methods. RNA from the mother and patient OI31 was analyzed as indicated.

(2), detection in this situation may be more certain using end-labeled probes with both senses assayed in separate tubes *(11)*. A possible strategy to cover the rare unreactive T base (Section 4.6 and ref. *12*) is to simultaneously use mutant and wild-type probes of both senses. This strategy has been used *(13)*, and would almost eliminate the chance of mutations being missed because of rare effects resulting from surrounding sequence, since each mutation has two chances of being detected directly. This strategy also allows prediction of the base change.

13. Applications of the method. New mutations causing inherited disease have been discovered using the CCM method in: (a) osteogenesis imperfecta (a type I collagen defect) *(15,16)*—indirect detection of an insertion has been reported in the collagen gene *(15)*; messenger RNA has been used in the studies of mutant collagen genes *(14–17)*; (b) orni-

thine transcarbamylase deficiency *(12)*; (c) hemophilia B *(18)*; and (d) tetrahydrobiopterin deficiency (dihydropteridine reductase deficiency) *(19)*. Polymorphisms near disease loci have also been detected *(12,15,17)*. In viral diseases, patterns of differences between viral strains have been obtained by hybridizing end-labeled DNA probes directly with viral RNA isolated directly from infected cells *(20)*.

References

1. Cotton, R. G. H. (1989) Detection of single base changes in nucleic acids *Biochem. J.* **263,** 1–10.
2. Cotton, R. G H , Rodrigues, N. R., and Campbell, R D. (1988) Reactivity of cytosine and thymine in single base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations *Proc Natl Acad Sci* **85,** 4397–4401
3. Cotton, R G. H. and Campbell, R D. (1989) Chemical reactivity of matched cytosine and thymine bases near mismatched and unmatched bases in a heteroduplex between DNA strands with multiple differences. *Nucleic Acids Res* **17,** 4223–4232
4. Maniatis, T , Fritsch, E. F., and Sambrook, J. (1989) *Molecular Cloning A Laboratory Manual,* 2nd Ed , Cold Springs Harbor Laboratory, Cold Spring Harbor, NY
5. Saiki, R. K., Scharf, S , Faloona, F., Mullis, K. G , Horn, G T Ehrlich, H A , and Arnheim, N. (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230,** 1350–1354
6. Shi, Y. and Tyler, B M (1989) Pyrrolidone, a noncontrolled substance, can replace piperidine for the chemical sequencing of DNA *Nucleic Acids Res* **17,** 3317
7. Schowalter, D B. and Sommer, S. S. (1989) The generation of radiolabeled DNA probes with polymerase chain reaction. *Anal. Biochem.* **177,** 90–94
8. Maxam, A. M and Gilbert, W. (1977) A new method for sequencing DNA. *Proc. Natl Acad Sci USA* **74,** 560–564.
9. Scharf, S J., Horn, G T , and Erlich, H. A. (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences *Science* **233,** 1076, 1077.
10. Dianzani, I , Forrest, S. F., Camaschella, C , Gottardi, E., and Cotton, R. G. H. Heterozygote identification by chemical cleavage of mismatch *Am J. Hum Genet* (in press)
11. Grompe, M , Muzny, D M., and Caskey, C T. (1989) Scanning detection of mutations in human ornithine transcarbamylase (OTC) by chemical mismatch cleavage *Proc Natl. Acad. Sci. USA* **86,** 5888–5892.
12. Forrest, S. M., Dahl, H H , Howells, D. W., Dianzani, I , and Cotton R G. H. (1991) Mutation detection in phenylketonuria using the chemical cleavage of mismatch method: Importance of using probes from both normal and patient samples. *Am J Hum. Genet* (in press).
13. Han, M , and Sterberg, P W (1990) let-60, a gene that specifies cell fates during *C elegans* vulval induction, encodes a *ras* protein *Cell* **63,** 921–931
14. Dahl, H-H M., Lamande, S. R , Cotton, R G. H , Cole, W G , and Bateman, J F (1990) A rapid chemical cleavage method for the detection and localization of base changes in RNA. *Proceeding of the UCLA symposia on molecular and cellular biology, Biotechnology and human genetic predisposition to disease,* vol. 126 Cantor, C R , Caskey, C T., Hood, L. E., Kamely, D , and Ommen, G. S , eds Alan R Liss, NY, pp 209–216

15. Bateman, J F., Lamande, S R , Dahl, H-H. M , Chan, D , Mascara, T., and Cole, W G. (1989) A frameshift mutation results in a truncated non-functional carboxy-terminal pro α 1(I) propeptide of type I collagen in osteogenesis imperfecta. *J. Biol. Chem.* **264,** 10960–10964

16. Lamande, S R , Dahl, H-H M., Cole, W. G., and Bateman, J. F. (1989) Characterization of point mutations in the collagen COLIA1 and COLIA2 genes causing lethal perinatal osteogenesis imperfecta. *J. Biol. Chem.* **264,** 15809–15812

17 Dahl, H-H. M., Lamande, S. R., Cotton, R. G. H., and Bateman, J F. (1989) Detection and localization of base changes in RNA using a chemical cleavage method. *Anal. Biochem* **183,** 263–268

18. Montandon, A. J , Green, P M , Gianneli, R., and Bentley, D. R (1989) Direct detection of point mutations by mismatch analysis: Application to haemophilia B. *Nucleic Acids Res.* **17,** 3347–3358

19 Howells, D W., Forrest, S M , Dahl, H-H. M and Cotton, R. G H. (1990) Insertion of an extra codon for threonine is a cause of dihydropteridine reductase deficiency *Am J Hum. Genet.* **47,** 279–285.

20 Cotton, R. G. H and Wright, P J (1989) Rapid chemical mapping of dengue virus variability using RNA isolated directly from cells. *J. Virol. Methods* **26,** 67–76

CHAPTER 6

Rapid Methods for Detection of Polymorphic Markers in Genomic DNA

David R. Bentley, Roland G. Roberts, and Jane Montandon

1. Introduction

The identification and use of restriction fragment length polymorphisms (RFLPs) *(1)* detected by Southern blotting *(2)*, using known or anonymous DNA probes, has provided the means for development of genetic markers that are distributed throughout the genome and that form the basis for linkage maps *(3,4)*. The approach used to identify new RFLPs with a given probe has been to analyze a number of genomic DNA samples from unrelated individuals, digested with a range of restriction enzymes. For example, Aldridge et al. *(5)* were able to detect RFLPs in five probes out of a total of 15 tested by assaying 23 restriction digests of three genomic DNA samples. This approach is inefficient, however, as it only detects sequence changes that alter the restriction sites being assayed, plus deletions or insertions that are large enough to alter restriction fragment sizes.

Rapid methods for detecting sequence variation in genomic DNA, such as RNase A cleavage of mismatched RNA:DNA hybrids *(6)*, denaturing gradient gel electrophoresis (DGGE) *(7,8)* of DNA heteroduplexes, and chemical mismatch detection *(9,10)* of DNA amplified by the polymerase chain reaction (PCR) *(11,12)*, can be used for more efficient detection of polymorphic markers in a given region as the sequence variation is detected, whether it lies in a restriction site or not. Of these three methods, chemical mismatch detection is potentially the most efficient, as it is able to detect all types of

From *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

sequence variation in genomic DNA *(13)*, whereas DGGE and RNase methods can detect approx 80–90% and 70%, respectively, of all sequence changes. The first part of this chapter describes one approach used to detect sequence variation by amplification and mismatch detection (AMD) analysis. In contrast to other approaches, AMD gives exact information on the position of a sequence change. Multiple changes are detectable within one fragment, as the cleavage at each mismatch is partial. The sequence differences are then characterized more fully by direct sequencing of the amplified product as described by Green and Giannelli in Chapter 3 of this volume.

In order to use previously detected sequence variations routinely as polymorphic markers in linkage studies and for genetic diagnosis, it is necessary that the assay used to detect the polymorphism be rapid and reliable. The use of PCR to amplify a specific region of the genome forms the basis for a rapid nonradioactive assay of polymorphisms *(11,14,15)* that have previously been characterized. In the second part of this chapter we describe different methods used to assay polymorphisms, which are either restriction site changes, length variations, or single-base substitutions that do not alter restriction sites or restriction fragment lengths. To detect this class of polymorphism, it has been necessary to develop new approaches that distinguish the two forms of such polymorphisms with accuracy and reliability. The availability of an increased set of rapid assays for polymorphisms that can be used as genetic markers will result in considerable improvements to the resolution of genetic maps and diagnosis of genetic diseases in man.

1.1. Characterization of Novel Polymorphisms by Amplification and Mismatch Detection

The method for AMD analysis *(10,16)* is outlined in Fig. 1. DNA from a known or reference sample is amplified by PCR with a selected pair of primers. The unknown or target DNA is amplified using the same primers in an independent reaction. The reference sample is purified to remove excess primers and nucleotides remaining at the end of the PCR, and either both (as shown in Fig. 1) or one of the strands are then end-labeled with ^{32}P The labeled DNA ("probe") is mixed with a 10–20-fold molar excess of unlabeled target, the sample is denatured at 100°C, and reannealed. If there is a sequence difference between the probe and target DNAs, a mixture of four duplexes is formed: two homoduplexes (a) and (d) and two heteroduplexes (b) and (c). The excess of target over probe minimizes the formation of the probe homoduplexes. This mixture is then divided into aliquots for treatment with either hydroxylamine, which modifies the C5=C6 double bonds, and thus destabilizes the pyrimidine ring of mispaired C residues *(17)*; or osmium tetroxide, which oxidizes the C5=C6 double bonds of the pyrimidine ring of mispaired T residues *(18)*. The DNA strands are then

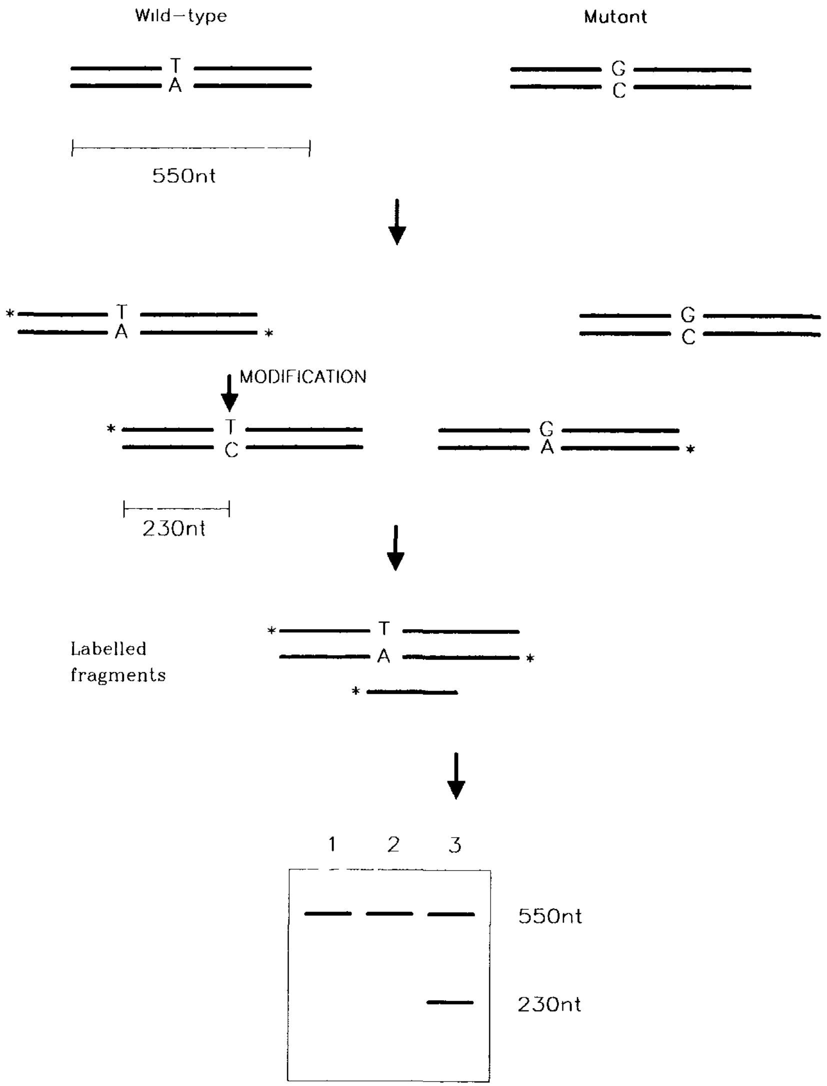


Fig. 1. Schematic diagram of procedure for mismatch detection. *Denotes radioactively labeled end; nt = nucleotides.

cleaved by piperidine at the sites of modification. The occurrence of a mismatch between probe and target DNA, which contains a C or a T in the probe sequence in the heteroduplex, therefore results in cleavage of the labeled probe. A shorter end-labeled fragment is generated, which can be detected following denaturing polyacrylamide gel electrophoresis. If both strands are assayed, potentially all sequence differences can be detected: a heteroduplex

containing an undetectable G or A at the mismatch site in one probe strand is complemented by the heteroduplex that contains the opposite probe strand, with a C or T, respectively, at the corresponding mismatch position.

2. Materials

2.1. Preparation of Amplified DNA

1. 10× PCR buffer: 670 m*M* Tris-HCl, pH 8.8, 166 m*M* $(NH_4)_2SO_4$, 67 m*M* $MgCl_2$.
2. Deoxynucleotide triphosphates (dNTPs): 5 m*M* with respect to each.
3. Oligonucleotide primers for forward and reverse reactions, each dissolved at 100 ng/µL.
4. Bovine serum albumin (BSA): 5 mg/mL.
5. β-Mercaptoethanol: 5% v/v.
6. *Taq* polymerase (e.g., AmpliTaq® from Cetus): 5 U/µL.
7. TE buffer: 10 m*M* Tris-HCl, 0.1 m*M* EDTA, pH 8.0.
8. Geneclean DNA purification kit: Bio 101.

2.2. Radiolabeling of Reference DNA

1. γ-^{32}P-ATP: 3000 Ci/mmol; 10 mCi/mL.
2. 10 × 5' labeling buffer: 0.5*M* Tris-HCl, 0.1*M* $MgCl_2$, 0.1*M* dithiothreitol (DTT), pH 7.6.
3. T4 polynucleotide kinase.
4. α-^{32}P-dNTP: 3000 Ci/mmol; 10 mCi/mL.
5. Klenow polymerase: 5 U/µL.

2.3. Mismatch Detection (See Note 1 for Safety Precautions)

1. 10× Hybridization buffer: 3*M* NaCl, 1*M* Tris-HCl, pH 8.0.
2. Mussel glycogen: 20 mg/mL in water.
3. C-reaction solution: Prepare a 4*M* solution of hydroxylamine hydrochloride (278 mg/mL in water) and add 0.3 vol of diethylamine to bring the pH to 6.0.
4. T-reaction solution: Prepare a solution containing 0.025% osmium tetroxide (we use a 4% stock solution from Aldrich Chemical Co. Ltd., Ellingham, Dorset, UK), 3% pyridine, 5 m*M* Tris-HCl, pH 8.0, 0.5 m*M* EDTA.
5. Piperidine: 10*M* Stock solution, diluted to 1*M* just before use by adding 100 µL stock under 0.9 mL water in a 1.5 mL microfuge tube on ice. Cap immediately to minimize evaporation.

6. Formamide dyes: 95% Deionized formamide, 10 m*M* EDTA, pH 8.0, 10 mg/mL xylene cyanol, 10 mg/mL bromophenol blue.
7. 6–8% Polyacrylamide gel: 1:20 Bisacrylamide:acrylamide, 7*M* urea, 40 cm × 20 cm × 0.3 mm gels in Tris/borate/EDTA buffer (TBE). 10× TBE contains 108 g Tris base, 55 g boric acid, and 40 mL 0.5*M* EDTA, pH 8.0, per liter.

2.4. Detection of RFLPs

1. 10× Restriction buffer and restriction enzymes: Commercially available.
2. 40% Acrylamide stock solution: 20:1 Acrylamide to bisacrylamide in deionized water.
3. 10× TBE buffer (*see* Section 2.3.7).
4. 10% Ammonium persulfate: Prepare from solid every 3–4 wk and store at 4°C.
5. Tetramethylethylenediamine (TEMED).
6. 1% w/v Ethidium bromide.

3. Methods

3.1. Preparation of Amplified DNA

In order to detect sequence variation in a given region of the genome, PCR primers are designed to enable amplification of the chosen region for study. Previous estimates indicate that the frequency of sequence changes in the genome is 1/300 bases (autosomes) or 1/1100 (X chromosome) *(5)*. AMD analysis of 1–2 kb of noncoding sequence is therefore likely to result in detection of polymorphisms. It is advisable to limit the length of the PCR to a maximum of about 1 kb. Our PCR method is given below.

1. Set up the following PCR:

	μL	
Reaction premix (per reaction)	5	10× PCR buffer
	5	5 m*M* dNTPs
	5	Primer 1 (100 ng/μL)
	5	Primer 2 (100 ng/μL)
	1.7	bovine serum albumin (BSA)
	0.7	β-mercaptoethanol (5%)
	0.6	*Taq* polymerase (AmpliTaq®; 5 U/μL)
	17	TE buffer
	40 μL	

2. Premix all components and add to genomic DNA (approx 10 μL; 50–100 ng). Heat reaction at 94°C for 4 min, then incubate for 30 cycles at

93°C, 1 min; 55–65°C (the optimum temperature depends on the primers) 1 min; 72°C, 2–7 min (depending on length of product; we use 7 min for product greater than about 700 bp, or alternatively, a 2–3-min initial extension time with an 8-s increment at each cycle). At the end of the reaction, incubate at 72°C for 5 min.

3. Analyze 5–10 µL of the reaction products on 0.7–1% agarose or 5% polyacrylamide minigels (run in 1× TBE buffer).
4. Purify the remainder using Geneclean (Bio 101) to remove primers and other reaction components of the PCR, and recover DNA in 10–20 µL TE.

3.2. Radiolabeling of Reference DNA Sample

1. The reference sample can be radioactively labeled by one of two methods. The 5' end of both strands can be labeled using T4 polynucleotide kinase and γ-^{32}P-ATP. Alternatively, if suitable restriction sites are present in the natural sequence or are incorporated into the ends of the PCR primers, the target DNA is restricted and the 3' ends of the strands labeled by incorporation of the appropriate α-^{32}P-dNTPs in a fill-in reaction, using the Klenow fragment of DNA polymerase I. In the latter case, it is possible to label one strand selectively if suitable restriction sites are available. For example, SalI and ClaI sites can be incorporated at the respective ends of the DNA fragment so that either can be digested and labeled separately.

   Method A: 5'end labeling. A ratio of (γ-^{32}P-ATP):(5' termini) of 1–2:1 is required in order to obtain maximum specific activity while keeping unincorporated ^{32}P-nucleotides to a minimum.

   Set up the following reaction:

µL	
5.0	DNA (100–500 ng total; 5–10 ng is required per mismatch reaction)
0.5	α-^{32}P-ATP
1.0	10× 5' buffer
0.5	T4 polynucleotide kinase
3.0	TE buffer
10 µL	

Incubate at 37°C for 15–30 min.

Method B: 3' end labeling. Digest the DNA with a large excess of the appropriate restriction enzyme (e.g., 5–10 U/0.5–1 µg DNA), and set up the following reaction:

μL	
5	DNA (digested) (100–500 ng total; 5–10 ng is required per mismatch reaction)
1	α-^{32}P-dNTP
1	TM buffer
0.5	Klenow polymerase
1.5	TE buffer
10 μL	

Incubate at 37°C for 10 min.

2. To monitor the success of probe labeling where both ends have been labeled, take an aliquot (1/10 or 1/20) of the probe, digest it with a restriction enzyme that cuts the fragment into at least two pieces, electrophorese the products on an acrylamide gel (this can be the same gel as that used for the subsequent mismatch analysis), and autoradiograph. The two bands observed should be of equal intensity.
3. Probes that have been labeled by either method may be used directly in the mismatch reaction. The degree of background obtained in the final autoradiograph, however, depends on the quality of the probe. If a high level of background bands is observed, this may be owing to the presence of short fragments in the probe preparation resulting from incomplete extension during amplification. To avoid this, it is advisable to optimize the PCR conditions (*see* Note 2). Alternatively, the probe may be purified by elution from a denaturing 4% polyacrylamide, 8*M* urea, thin (0.3-mm) gel. The gel slice containing the probe is shaken in 0.4 mL 2*M* ammonium acetate at 37°C overnight, and the DNA recovered from the aqueous fraction by ethanol precipitation.

3.3. Mismatch Detection

3.3.1 Preparation of Hybrids

1. Prepare hybrids by mixing 5–10 ng of labeled probe (e.g., 10 ng of a 1-kb fragment, approx 0 01 pmol) with a 10–20-fold molar excess of unlabeled amplified target.

μL	
10	labeled probe (5–10 ng)
2	target (100–200 ng)
2	10× hybridization buffer
6	TE
20 μL	

Cover reaction in 0.5 mL Eppendorf tube with paraffin oil. Incubate at 100°C for 5 min, then bring directly to 65°C and incubate for 5–18 h.

2. In order to test for the formation of duplexes in the hybridization step, two aliquots of the annealing reaction can be analyzed on a denaturing gel. One sample is denatured before loading; the other is loaded in glycerol dyes, and runs in native form. Also, both denatured and native markers should be loaded. The successful formation of hybrids is indicated by the presence of a band at the appropriate size in the nondenatured sample. Unhybridized probe in the hybridization reaction will comigrate with the band in the denatured sample.
3. Transfer the aqueous phase to siliconized 1.5 mL Eppendorf tubes, add 3 µL mussel glycogen, and recover sample by ethanol precipitation. Resuspend in 14 µL and divide into two equal aliquots, for treatment with hydroxylamine or osmium tetroxide, respectively.

3.3.2. Chemical Modification

1. C-reaction *(17)*: Add 20 µL of C-reaction solution (final concentration of hydroxylamine: 2.3*M*) to each sample and incubate at 37°C for 2 h.
2. T-reaction *(17,19)*: Add 18 µL of the T-reaction solution to each sample and incubate at 37°C for 2 h.
3. Cleavage: To each sample, add 100 µL 3*M* sodium acetate, 0.1 m*M* EDTA, 300 µL ethanol, and precipitate. Recover precipitate, wash pellet twice in 70% ethanol, and air-dry. Resuspend by vortexing in 50 µL of freshly diluted 1*M* piperidine. Cap tubes tightly and incubate at 90°C for 30 min: we use a dry-block and keep the tubes sealed by placing an aluminum block over the caps. Recover samples by ethanol precipitation and wash pellets in 70% ethanol.
4. Electrophoresis: Resuspend samples in 5 µL TE plus 2 µL formamide dyes. Boil the samples 3 min to denature DNA and load on 6–8% polyacrylamide gels Electrophorese at 30 mA for approx 2 h and autoradiograph.

In the example shown in Fig. 2, a 1-kb region of the 3' untranslated region of the dystrophin gene was subjected to AMD analysis. DNA from five unrelated females was analyzed using the DNA from one of them (sample 5) as a probe. Using osmium tetroxide modification, bands of approx 750 nucleotides (nt) and 650 nt, plus a ladder of 10 bands at 220–260 nt were present in all samples; and sample 3 had additional bands at 850 nt, plus 150 nt (3 bands). This result indicated the presence of a number of sequence polymorphisms within the region.

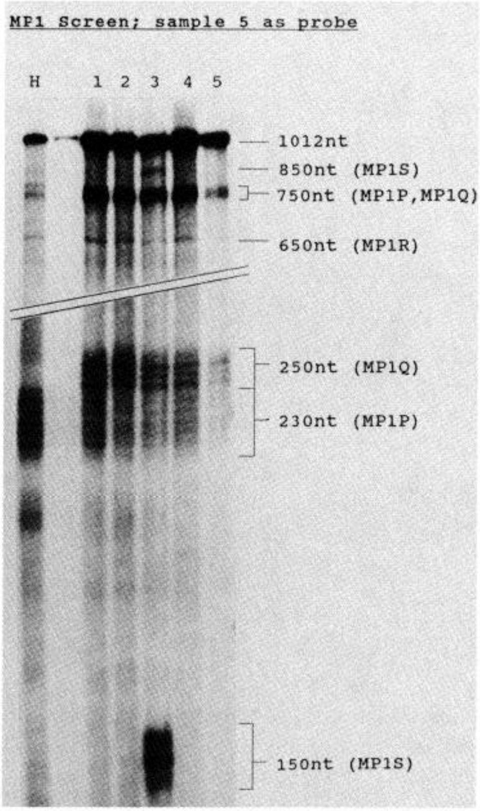


Fig. 2. AMD analysis of 1 kb of the 3' untranslated region of the dystrophin gene using osmium tetroxide (lanes 1–5) or hydroxylamine (lane H). Bands corresponding to each of the polymorphisms are bracketed, and each band is accounted for in subsequent sequence analysis. For example, each of the six bands in the 230-nucleotide (nt) region for MPlP is the result of one of the short tandem repeats looping out in a (6-copy): (5-copy) heteroduplex.

Subsequent sequence analysis of the amplified product from selected samples showed the presence of four polymorphisms: MP1P, Q, R, and S. Figure 3, for example, shows sequence determination of the different forms of the MP1P polymorphism. The MP1P polymorphism is a length variation, in which the two major alleles are five and six copies, respectively, of the tetranucleotide repeat TTGA. The left-hand sequence is of a 6-copy allele, the right-hand sequence is of a 5-copy allele, and the central sequence is of a heterozygote, where the two sequences are superimposed.

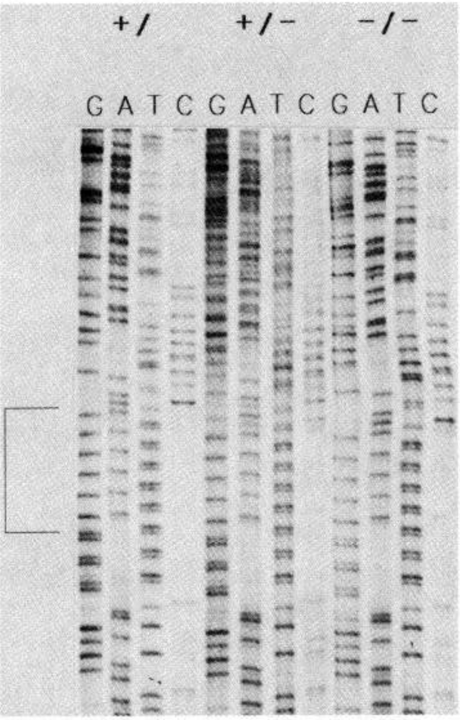


Fig. 3. Direct sequence analysis of the MPIP polymorphism. Left: sequence of DNA from an individual who is homozygous for the 6-copy allele (6 copies of the TTGA repeat are shown bracketed). Right: sequence of a 5-copy sample. Center: sequence of a heterozygous sample.

3.4. Rapid Assays for Detection of Known Polymorphisms

Sequence analysis around the site of the polymorphism permits the design of primers for specific, rapid, PCR-based assay of each polymorphism. This is necessary prior to introduction of the marker for diagnosis, or linkage analysis, in which large numbers of samples are to be tested for the polymorphic marker. The types of sequence variation that are used as polymorphic markers can be classified into three groups: (i) RFLPs: these may be caused by either a single-base substitution or an insertion or deletion at the restriction enzyme recognition sequence; (ii) single-base substitutions that do not lie in a restriction site; and (iii) length polymorphisms that arise from insertion or deletion of one or more nucleotides in one allele compared to the other, and which do not lie in a restriction site.

3.4.1. Detection of RFLPs by Digestion of PCR Products

1. Perform PCR using primers that flank the polymorphic restriction site.
2. Make a premix for the required restriction enzyme as follows. Volumes are given per reaction and for 10 reactions. In the latter case, sufficient premix is made for 12 reactions to allow for losses during aliquoting. It is assumed that the PCR volume is 25 µL.

Premix:	for 1 reaction	for 10 reactions
10× restriction buffer	3 µL	36 µL
Restriction enzyme	1–2 µL (10 U)	12–24 µL (10 U)
	4–5 µL	48–60 µL

Add appropriate volume (4–5 µL) to the side of tube containing PCR products. Flick-spin, vortex, and flick-spin again. Incubate tubes at the appropriate temperature for 60–120 min. Analyze 5–12.5 µL of each sample on a 3% Nusieve agarose minigel, or a 5% or 12% polyacrylamide minigel. Polyacrylamide minigels generally give better resolution (sharper bands) than Nusieve gels, but considerable quenching of the fluorescent signal from the ethidium bromide-stained bands occurs in polyacrylamide (approx fourfold reduction in intensity of bands), necessitating the loading of larger amounts of sample. Gels are run in 1× TBE.

3. Prepare polyacrylamide minigels as follows:

Gel mixes:	5% stock	12% stock
40% (20:1 acrylamide: bisacrylamide, deionized)	62.5 mL	150 mL
10× TBE	50 mL	50 mL
1% ethidium bromide	25 µL	25 µL.

Make up to 500 mL in each case with deionized water.

To 40–50 mL of stock mix in a 100 mL conical flask, add 400 µL of 10% ammonium persulfate (make up 10% stock from solid every 3–4 wk and store at 4°C) and 140 µL TEMED. Mix by swirling, and pour gel. Allow to set for at least 5 min. Pour off unset acrylamide and rinse with 1× TBE. Remove segments of gel that may have leaked into electrode chambers. (This occurs in some minigel boxes that use slide-in partitions to retain the gel while it sets.) Cover with 50 mL (approx) 1× TBE containing 2.5 µL of 1% ethidium bromide, and electrophorese at 90 mA for 1–2 h as required.

Figure 4 shows the results of analysis of a family with a case of Becker muscular dystrophy (BMD) using the pERT 87-15/XmnI polymorphism. In

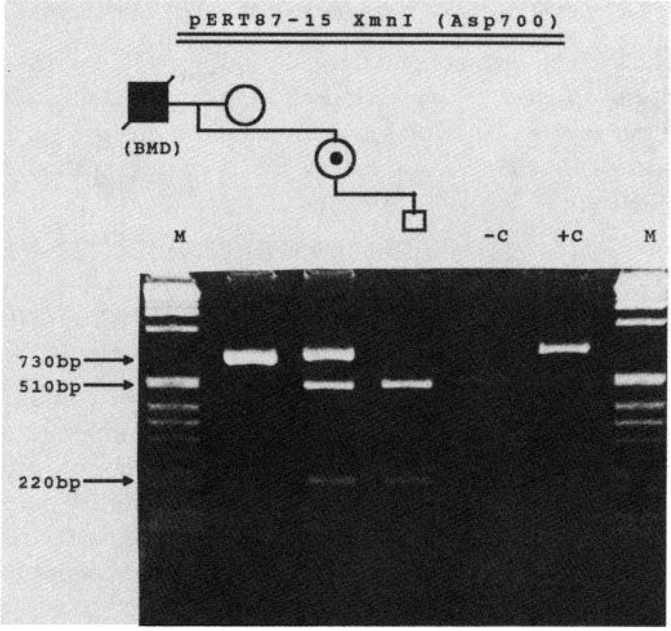


Fig. 4. PCR-based assay of an RFLP in the dystrophin gene (pERT87-15/XmnI). The analysis was carried out using the isoschizomer Asp700. In this family, the XmnI+ allele is detected by the appearance of the 510 bp and 220 bp bands after digestion of the PCR product. The fetus (diagnosed by amplification of DNA from a chorion villus sample) is hemizygous for the XmnI+ allele. His mother is heterozygous for the XmnI polymorphism: the XmnI-allele of the mother is grandmaternal in origin, and the chromosome with the XmnI+ allele by inference is the high-risk grandpaternal chromosome that carries the Becker muscular dystrophy mutation.

this case, the polymorphism was analyzed using Asp700, which is an isoschizomer of XmnI. The grandmaternal sample (lane 2) is homozygous for absence of the XmnI site (–/–). Her daughter (lane 3) is heterozygous (–/+), and the (+) allele is therefore carried on the grandpaternal chromosome, which also carries the BMD mutation. The male fetus has inherited the grandpaternal chromosome and therefore has a high risk of being affected with BMD.

3.4.2. Detection of Length Polymorphisms by Electrophoretic Separation of PCR Products

Polyacrylamide or Nusieve minigels can be used to resolve length variations as small as 3 bp, given that the initial PCR product is no more than 60–80 bp in length (*see,* for example, ref. *20*). In the case of MPlP, the initial

products were 82 bp and 78 bp. To improve resolution of the two forms, the PCR products are digested with AluI, which cuts at an invariant site to release fragments of 60 bp or 56 bp from the two forms, respectively. These are resolved readily on a 12% polyacrylamide minigel as shown in Fig. 5(A).

A common feature of the analysis of heterozygotes of length polymorphisms is the appearance of an additional higher band (*see* lanes 2, 3, and 4 in Fig. 5A). This is presumed to represent a heteroduplex formed between two DNA strands of different lengths. The single-stranded loop in the longer strand is sufficient to cause marked retardation on the native acrylamide gel.

Figure 5B shows another example of a length polymorphism, detected directly by analysis of the PCR products on a 12% acrylamide minigel. The two alleles are represented by 67- or 63-bp bands (further details of interpretation are given in figure legend).

3.4.3. Detection of Single-Base Substitutions that Do Not Alter Restriction Sites

The detection of single-base substitutions that cannot be assayed as RFLPs has been achieved by selective hybridization of radioactively labeled allele-specific oligonucleotide (ASO) probes to dot blots of amplified target (*see* ref. *21* and Chapters 7 and 24, this volume). The conditions of hybridization are optimized such that a probe hybridizes only to the exactly matched target sequence, and not the target sequence with a single mismatched base. This method is applicable to the detection of all single-base substitutions, but suffers the disadvantage that it is a radioactive procedure, and includes a hybridization step.

Two alternative procedures have been described in which the single-base difference between two alleles is used to determine specific priming of an appropriate oligonucleotide in the PCR. In the amplification refractory mutation system (ARMS) (*see* ref. *22* and Chapter 8, this volume), two primers (1 and 2 in Fig. 6A) are synthesized with sequences that are identical except for one base, which is specific to either allele of the sequence variation. In order to assay for the presence of both alleles in a DNA sample, two PCRs are performed, one with each primer plus a universal primer (primer 3). PCR product is only generated if the primer present exactly matches the template. Thus, PCR only occurs with primer 1 on allele Q (Fig. 6B) and with primer 2 on allele q (Fig. 6C). A representation of the gel is shown in Fig. 6D. If the target DNA sample is heterozygous for alleles 1 and 2, PCR will occur in both reactions. The position of the sequence differences in the allele-specific primers is critical. Newton et al. *(22)* reported that optimal positions were 1, 2, or 3 nucleotides from the 3' end. (In Fig. 6, the sequence difference is shown one nucleotide away from the 3' end.) A similar approach was adopted by Li et al. *(23)*, in which the sequence difference of the allele-specific primer

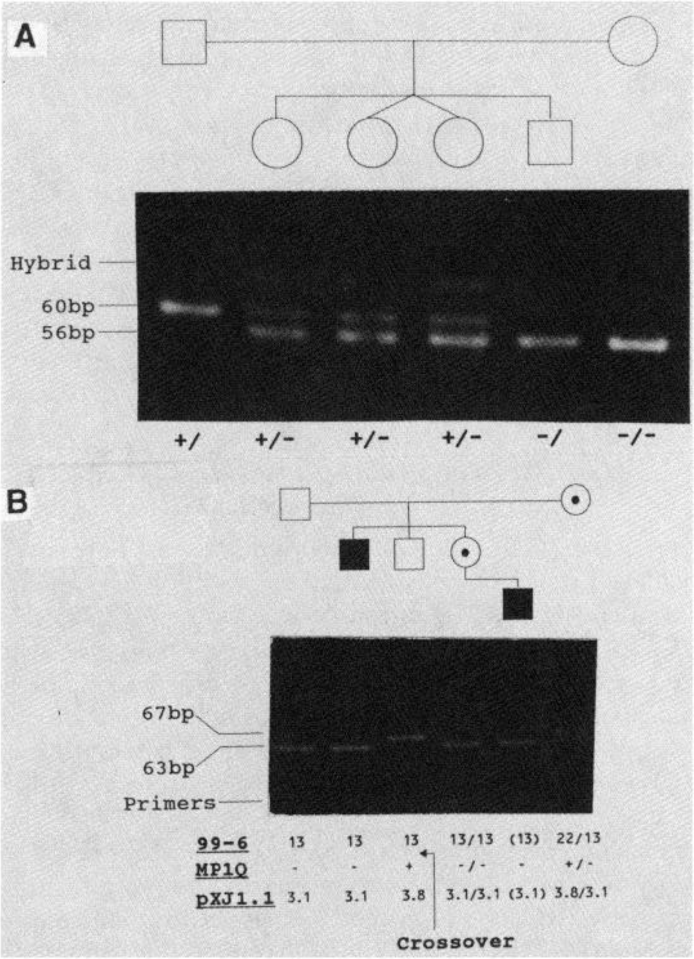


Fig. 5. Assay of length polymorphisms by PCR. **(A)** Segregation of the MPlP polymorphism: + and – denote alleles with 6 or 5 copies of the TTGA repeat and are detected by the appearance of the 60- bp or 56-bp bands, respectively, following digestion of the initial PCR product with AluI, which cuts at an invariant site. The additional higher band seen in the heterozygous samples is a heteroduplex, in which the single-stranded loop retards the fragment on the gel. **(B)** Segregation of the MP1Q polymorphism in a family with Duchenne muscular dystrophy. + and – denote alleles with 4 and 3 copies of the GTAA repeat and are detected by the appearance of 67-bp and 63-bp bands, respectively. The markers are in the order pXJ1.1 (intragenic), MP1Q (3' end) and 99–6 (3' flanking). Analysis of DNA from members of this pedigree with 99–6 and pXJ1.1 demonstrated that a crossover had occurred in the normal male (third lane) as he had inherited the 13 kb band detected with 99–6, but also the 3. 8 kb band detected by pXJ1.1. Analysis with MP1Q demonstrated that this individual also had the + allele. Thus, the crossover is between 99–6 and MP1Q, as indicated by the arrows.

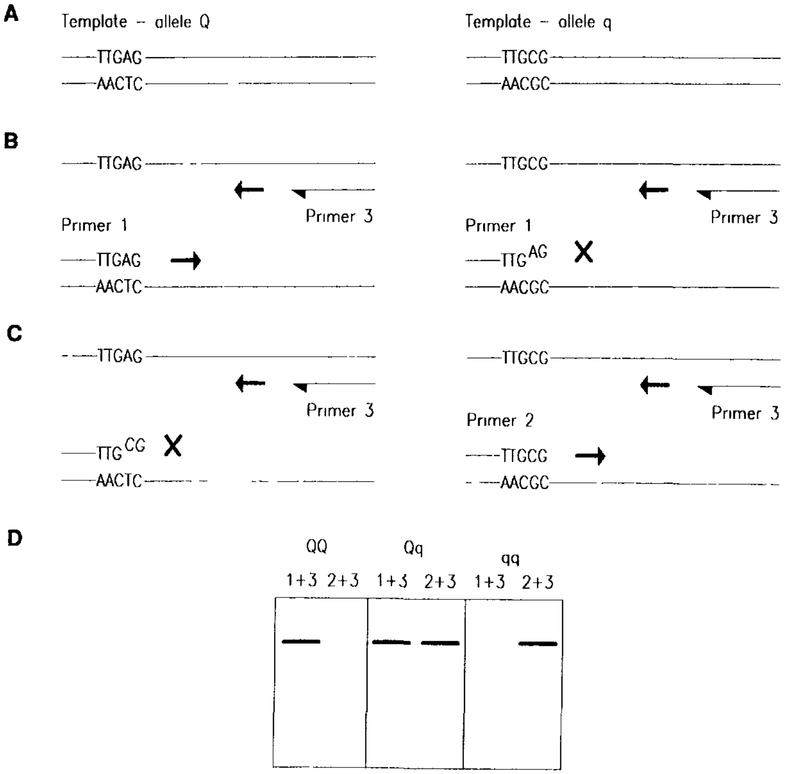


Fig. 6 Scheme for detection of single-base changes by ARMS. (A) Sequences represent two hypothetical alleles Q and q that differ by a single base (A vs C in the top strand). (B) Priming of each template with primers 1 and 3; primer 1 matches the sequence of allele Q and successful priming occurs, whereas no priming occurs with primer 1 on allele q. (C) Priming of each template with primers 2 and 3; primer 2 matches the sequence of allele q and successful priming occurs, whereas no priming occurs with primer 2 on allele Q. (D) Expected gel pattern. Genotypes of DNA samples used (QQ, Qq, qq), and primers used (1 + 3 or 2 + 3) are shown above the gel.

was at the 3' end of each primer. Specific priming was achieved by reducing the dNTP concentration to 4 μ*M*.

In the competitive oligonucleotide priming (COP) system, reported by Gibbs et al. *(24)*, the sequence difference that distinguishes the allele-specific primers is in the middle region of each oligonucleotide. For the assay, one PCR is carried out using all three primers. The two allele-specific primers thus compete for the appropriate target sequence, the correct match between primer and target providing in the order of 100-fold more efficient priming

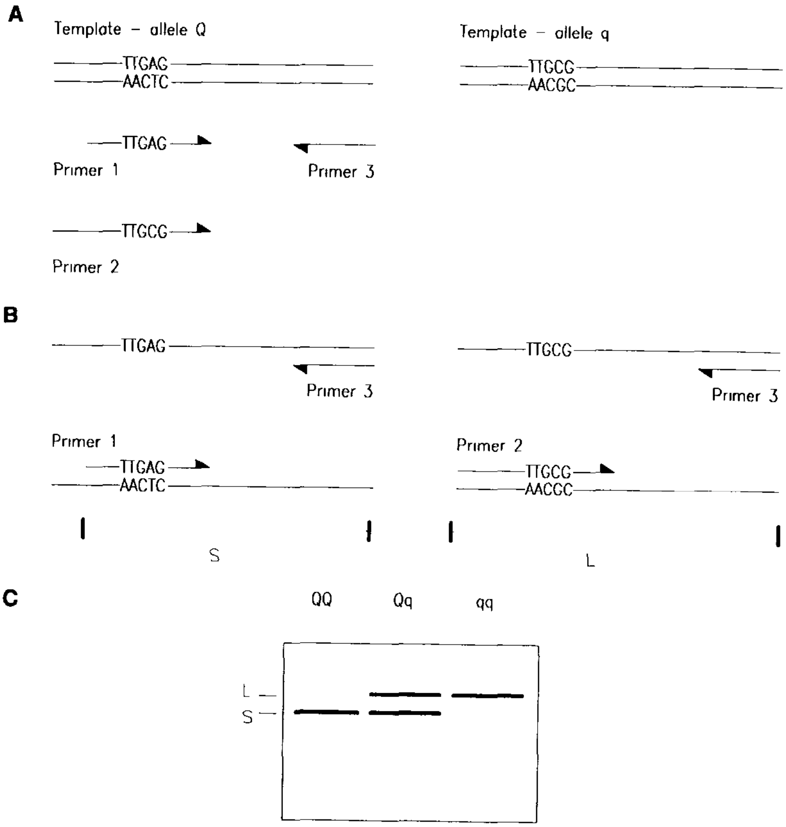


Fig. 7. Scheme for detection of single-base changes by COP. **(A)** Sequences represent two hypothetical alleles that differ by a single base (A vs C in the top strand). **(B)** A PCR is performed with all three primers present. Primer 1 competes successfully for the template of allele Q while primer 2 remains dissociated from it. Primer 2 competes successfully for the template of allele q while primer 1 remains dissociated from it. The two products are distinguished by the greater length of primer 2, which results in the formation of a longer PCR product with primer 2 than with primer 1. Products are denoted by dotted lines and, L (long) or S (short). **(C)** Expected gel pattern. Genotypes are shown above the gel diagram.

than a mismatched primer/template combination. The two products are distinguished either by a difference in length between the two allele-specific primers (Fig. 7A–C), or by the use of different fluorescent dyes attached to each primer. Use of the latter system has been reported for the detection of the sickle cell mutation by Chehab and Kan *(25)*.

4. Notes

1. Hydroxylamine, osmium tetroxide, and piperidine are toxic. Solutions should be prepared and used in a fume hood.
2. A low yield of PCR product may be improved by one or more of the following steps:
   a. Increasing the extension time by 1–5 min.
   b. Checking or replacing buffer components and dNTP stocks.
   c. Using a different batch of enzyme.
   d. Improving the quality of the template DNA, by, for example, phenol/chloroform extraction and ethanol precipitation.
   e. Reducing the amount of template DNA in the PCR, which is often an alternative to d.

References

1 Botstein, D., White, R., L , Skolnick, M , and Davis, W , D (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms *Am J Hum. Genet.* **32,** 314–331.

2. Southern, E M (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis , *J. Mol. Biol* **98,** 503–517.

3 White, R , Leppert, M , Bishop, D. T., Barker, D., Berkowitz, J., Brown, C , Callahan, P , Holm, T , and Jerominski, L (1985) Construction of linkage maps with DNA markers for human chromosomes *Nature* **313,** 101–105.

4. Donis-Keller, H , Green, P., Helms, C , Cartinour, S , Weiffenbach, B., Stephens, K., Keith, T P., Bowden, D W., Smith, D. R., Lander, E. S., Bostein, D , Akots, G., Rediker, K. S., Sravius, T , Brown, V A , Rising, M B., Parker, C , Powers, J A , Walt, D. E., Kauffman, E. R., Bricker, A , Phipps, P., Muller-Kahle, M., Fulton, T. R., Ng, S , Schumm, J W., Braman, J. C , Knowlton, R. G., Barker, D. F Crooks, S. M., Lincoln, S. E., Daly, M. J., and Abrahamson, J (1987) A genetic linkage map of the human genome. *Cell* **51,** 319–337

5. Aldridge, J , Kunkel, L , Bruns, G., Tantravahi, U., Lalande, M., Brewster, T , Moreau, E., Wilson, M., Bromley, W , Roderick, T., and Latt, S. A. (1984) A strategy to reveal high frequency RFLPs along the human X chromosome *Am. J. Hum Genet* **36,** 546–564.

6 Myers, R. M., Larin, Z , and Maniatis, T. (1985) Detection of single base substitutions by ribonuclease cleavage of mismatches in RNA·DNA duplexes *Science* **230,** 1242–1246

7. Fischer, S G and Lerman, L S (1983) DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels. Correspondence with melting theory *Proc. Natl. Acad. Sci. USA* **80,** 1579–1583

8. Sheffield, V C , Cox, D R , Lerman, L. S., and Myers, R M. (1989) Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc Natl Acad Sci USA* **86,** 232–236.

9 Cotton, R. G. H., Rodrigues, N., and Campbell, R. D. (1989) Reactivity of cytosine and

thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad Sci. USA* **85**, 4397–4401.

10. Montandon, A. J , Green, P. M., Giannelli, F. B., and Bentley, D. R (1989) Direct detection of point mutations by mismatch analysis: Application to haemophilia B. *Nucleic Acids Res.* **17**, 3347–3358.
11. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, H. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science* **230**, 1350–1354.
12 Saiki, R K., Gelfand, D H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.
13. Cotton, R. G H. and Campbell, R. D (1989) Chemical reactivity of matched cytosine and thymine bases near mismatched and unmatched bases in a heteroduplex between DNA strands with multiple differences *Nucleic Acids Res.* **17**, 4223–4233
14 Chehab, F. F , Doherty, M., Cai, S , Kan, Y. W., Cooper, S , and Rubin, E M. (1987) Detection of sickle cell anaemia and thalassaemia. *Nature* **329**, 293,294
15. Kogan, S. C., Doherty, M , and Gitschier, J. (1987) An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences Application to haemophilia A *N. Engl J Med.* **317**, 985–990.
16. Roberts, R. G., Montandon, A. J , Bobrow, M., and Bentley, D. R (1989) Detection of novel genetic markers by mismatch analysis *Nucleic Acids Res* **17**, 5961–5971
17 Johnston, B. H. and Rich, A (1985) Chemical probes of DNA information:L detection of Z-DNA at nucleotide resolution. *Cell* **42**, 713–724
18. Friedmann, T. and Brown, D. M (1978) Base-specific reactions useful for DNA sequencing. *Nucleic Acids Res.* **5**, 615–622.
19 McClellan, J. A., Palecek, E , and Lilley, D. M. (1986) $(A\text{-}T)_n$ tracts embedded in random sequence DNA — formation of a structure which is chemically reactive and torsionally deformable. *Nucleic Acids Res* **14**, 9291–9309.
20. Mathew, C. G., Roberts, R. G., Harris, A., Bentley, D. R., and Bobrow, M. (1989) Rapid screening for the deltaF508 deletion in cystic fibrosis. *Lancet* **ii**, 1346.
21. Saiki, R. K., Bugawan, T. L., Horn, G .T., Mullis, K. B., and Erlich, H. A. (1986) Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. *Nature* **324**, 163–166.
22. Newton, C. R., Graham, A , Heptinstall, L E., Powell, S J., Summers, C , Kalshekar, N , Smith, J. C., and Markham, A. F. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503–2516
23. Li, H , Cui, X., and Arnheim, N. (1990) Direct electrophoretic detection of the allelic state of single DNA molecules in human sperm by using the polymerase chain reaction. *Proc. Natl. Acad Sci. USA* **87**, 4580–4584.
24. Gibbs, R A., Nguyen, P. N , and Caskey, C. T (1989) Detection of single DNA base differences by competitive oligonucleotide priming. *Nucleic Acids Res.* **17**, 2437–2448.
25. Chehab, F. F and Kan, Y. W, (1989) Detection of specific DNA sequences by fluorescence amplification. A color complementation assay. *Proc. Natl. Acad. Sci. USA* **86**, 9178–9182.

CHAPTER 7

The Analysis of Point Mutations Using Synthetic Oligonucleotide Probes

Christine J. Farr

1. Introduction

This chapter deals with the use of synthetic oligonucleotides as probes for the detection of allelic sequence variation and point mutations by use of a dot-blot procedure in DNA enzymatically amplified in vitro (polymerase chain reaction). The approach is based on the principle that a fully matched hybrid formed between the target sequence and an oligonucleotide probe is thermally more stable than a single base-pair (bp) mismatch hybrid. The differential hybridization under stringent conditions of short, radioactively labeled oligonucleotide probes was first introduced in the late 1970s, and is a powerful diagnostic tool *(1)*. However, because of the complexity of human genomic DNA and inefficiencies in hybridization that are inherent in short oligonucleotide probes, microgram quantities of sample DNA, gel electrophoresis, and highly radioactive DNA probes were required. Recently, a modification of this approach has been described that uses the polymerase chain reaction (PCR) to selectively amplify the DNA region of interest *(2)*. As a result of this substantial reduction in complexity, a simplified "dot-blot" detection format is possible that requires only nanogram amounts of sample DNA and moderately radioactive probes.

In the design of oligonucleotide probes, several aspects of the DNA sequence need to be considered, e.g., length, G+C content, and self-complementarity. The oligonucleotides should be long enough to be unique within the target genome (17 nucleotides for the mammalian genome) and short

From: *Methods in Molecular Biology, Vol 9· Protocols in Human Molecular Genetics*
Edited by: C Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

enough to allow the detection of mismatches under the conditions used. Typically, oligonucleotides used for probing mammalian DNA are 19–21 bases long. Self-complementarity of oligonucleotides creates problems for oligonucleotide purification and may affect the efficiency of hybridization to complementary DNA, and thus, should be avoided whenever possible The base composition is also important in duplex stability. For perfectly matched duplexes between 11–23 bases long in 1*M* Na^+, the dissociation temperature (T_d, temperature at which 50% of the duplex is dissociated) can be derived.

$$T_d = 2° \text{ (number of A+T residues) } + 4° \text{ (number of G + C residues)}$$

This relationship is only valid within these ranges of length and is meant to serve as a guide *(3)*.

Recently, a technique that uses tetramethylammonium chloride (M_4NCl) to eliminate the dependence of T_d on the G.C content of the probe has been described *(4)*. M_4NCl binds selectively to A.T bp, eliminating the preferential melting of A.T vs G.C bp, allowing the stringency of the hybridization to be controlled as a function of length only. Moreover, 19–20-base oligonucleotides melt over a smaller temperature range in solvents containing M_4NCl than in solvents containing sodium salts *(5)*.

The effect of a single bp mismatch on the stability of oligonucleotide–DNA hybrids will depend both on the type of mismatch and on its position relative to the end of the hybrid. Least stable will be a duplex with a central mismatch.

When used for detecting mutations, oligonucleotides are used in pairs (or as a panel of probes); one member is homologous to the normal sequence and the other is homologous to the mutated sequence. It is necessary to establish hybridization and washing conditions before embarking on such an analysis. The method described below uses tetramethylammonium chloride and 20-base oligonucleotide probes.

2. Materials

1. Oligonucleotides—gel or HPLC-size purified.
2. TE buffer: 10 m*M* Tris-HCl, pH 8.0; 1m*M* EDTA.
3. 1*M* Tris-HCl, pH 7.2 and pH 8.
4. 500 m*M* EDTA, pH 8.
5. 10*N* NaOH.
6. 20% SDS.
7. 1*M* $MgCl_2$.
8. 1*M* DTT (dithiothreitol).
9. 10× Kinase buffer: 250 m*M* Tris-HCl, pH 8; 100 m*M* $MgCl_2$; 100 m*M* DTT.

10. [γ-^{32}P]ATP specific activity (SA) > 5000 Ci/mmol (Amersham PB 10218).
11. 100% Trichloroacetic Acid (TCA).
12. Carrier DNA—500 µg/mL of sheared, denatured salmon sperm DNA (Sigma type III sodium salt) in TE.
13. 20× SSPE: 200 m*M* Sodium phosphate, pH 7.2; 3.6*M* NaCl; 200 m*M* EDTA.
14. 5*M* Tetramethylammonium chloride (M_4NCl) aqueous (Aldrich). The actual molar concentration (C) must be determined from the refractive index (N) by the formula C = (N – 1.331)/0.018.
15. 100× Denhardt's solution: 2% Ficoll (type 400, Pharmacia); 2% polyvinylpryrrolidone; 2% bovine serum albumin (fraction V).
16. Tetramethylammonium chloride hybridization buffer: 3*M* M_4NCl; 50 m*M* Tris-HCl, pH 7.2; 2 m*M* EDTA; 0.3% SDS; denatured, sonicated salmon sperm DNA (100 µg/mL); 5× Denhardt's solution.
17. Deionized Formamide—stir with Dowex XG8 mixed-bed resin for 1 h and filter twice through Whatman No. 1 paper.
18. Bio-Gel P4 Fine (Bio-Rad); a 10% slurry in TE.
19. 1 mL Disposable syringe and glass wool.
20. Whatman GF/C glass fiber filter paper, vacuum driven filtration manifold.
21. Dot-blot (or slot-blot) apparatus (Bio-Rad).
22. Nylon filter, e.g., Genetran-45 (Plasco) or Hybond N^+ (Amersham).
23. UV (254 nm) transilluminator for fixing DNA to filters. Caution: Shield UV source and wear protective goggles (alternatively, an 80°C oven).
24. Perspex shield, safety glasses, and disposable gloves for work with ^{32}P.

3. Methods

3.1. Determination of Oligonucleotide Concentration

Since the base composition of different synthetic sequences can vary widely, it is necessary to calculate the molar extinction coefficient (MEC) for the particular sequence in order to determine the precise concentration. The MEC at 260 nm (pH 8) is calculated by summing the contribution of each base: G, 12010; A, 15200; T, 8400, and C, 7050.

$$\text{Concentration (mol/L)} = \frac{OD_{260}}{MEC}$$

As a rough guide, 1 OD_{260} is approx 33 µg/mL. Solutions of oligonucleotides should be made in TE and stored frozen.

3.2. Oligonucleotide Probe Preparation

Oligonucleotides synthesized by the phosphotriester or the phosphoramidite approaches contain a free 5'–OH.

3.2.1. 5'-End ^{32}P-Labeling (Kinase Reaction)

This method is sufficient for most routine applications of oligonucleotide hybridization. It involves the transfer of [^{32}P]phosphate from [γ-^{32}P]ATP using T4 polynucleotide kinase.

1. Prepare in 40 µL of 1× kinase buffer: 20 pmol oligonucleotide, 40 pmol [γ-^{32}P]ATP (SA > 5000 Ci/mmol), and 10 U of T4 polynucleotide kinase.
2. Incubate the reaction at 37°C for 30–60 min. Stop the reaction by the addition of 60 µL of 25 m*M* EDTA.

At best, phosphorylation results in the incorporation of one atom of ^{32}P per molecule of oligomer. Thus, the SA achieved is limited by that of the [γ-^{32}P] ATP used (*see* Note 8).

3.3. Probe Purification

The extent to which radiolabeled oligonucleotide probes need to be purified varies according to their eventual use. For the detection of point mutations in DNA in which the target sequence has been amplified in vitro (PCR), the SA is less critical and it is not necessary, routinely, to separate the radiolabeled ("hot") oligonucleotide from contaminating unlabeled ("cold") oligonucleotide. A clean, simple separation of the oligonucleotide from the unincorporated [γ-^{32}P] ATP can be achieved by spin dialysis column chromatography using Bio-Gel P4 Fine as the gel filtration medium (*see* Note 5).

1. Plug the bottom of a 1-mL syringe with some glass wool and place it in a 15-mL Falcon tube.
2. Fill the syringe with Bio-Gel slurry and spin for 10–30 s at 1500–2000 rpm in a swinging bucket bench-top centrifuge.
3. Add more slurry and spin again. Continue until the bed vol is approx 1 mL.
4. Load 100 µL TE and spin for 2 min.
5. Transfer the packed syringe to a new Falcon tube containing an open screw-capped 1.5-mL Eppendorf tube. Adjust the vol of the sample to 100 µL and load it onto the column. Spin for 2 min and collect the flowthrough. If required, additional fractions can be collected by the addition of further 100-µL aliquots of TE.

3.4. TCA Precipitation of Oligomers Following Phosphorylation

1. Remove 2 µL of the kinase reaction (1/50) and dilute into 198-µL carrier DNA.
2. Mix 10 µL of this dilution with 100 µL of additional carrier and add 5 mL of ice-cold 10% TCA. Incubate on ice for 30 min.
3. Filter through Whatman GF/C filter, rinse 2–3 times with 10 mL ice-cold 10% TCA, and once with 10 mL ice-cold 100% EtOH.
4. Dry-filter and count on scintillation counter.

Specific activity of oligonucleotide probe:

$$\frac{\text{TCA counts} \times \text{dilution factor}}{2.22 \times 10^6} = \text{total incorporation } \mu\text{Ci}$$

$$\frac{\text{total incorporation } \mu\text{Ci}}{\text{pmol oligonucleotide}} = \text{specific activity } (\mu\text{Ci/pmol})$$

The specific activity of the phosphorylated oligomer is usually 4 µCi/pmol (or approx 1.5×10^9 dpm/µg) when using 5000 Ci/mmol [γ-^{32}P] ATP (i.e., approx 80% of the specific activity of the label). To determine the total counts, take 10 µL of the first dilution and multiply by 1000.

3.5. Dot-Blot Screening Procedure

Allelic sequence variation in genomic DNA in which the target sequence has been amplified in vitro can be analyzed using a simple "dot blot" for probe hybridization (*see* Note 6).

3.5.1. Transfer of DNA to Membrane

1. Subject genomic DNA (1 µg) to 30 cycles of standard PCR amplification (*see* Chapter 1). Adjust 10 µL of the final reaction mixture (equivalent to 100 ng of the original DNA), to 0.4*M* NaOH; 25 m*M* EDTA in a 200-µL vol.
2. Wet a Genetran-45 nylon filter (Plasco) in water and apply the denatured sample under vacuum with a Bio-Dot apparatus (Bio-Rad). Caution: Always wear gloves when handling the membrane.
3. Wash wells with 400 µL 20x SSPE using a multipipet.
4. Rinse the filters in 20x SSPE, blot with filter paper, and allow to air-dry.
5. Fix the DNA by baking (2 h at 80°C) or by UV illumination (wrap in Saran Wrap™ and irradiate, DNA-side down, on a standard UV transilluminator for 2–5 min).
6. Replicate filters can be prepared.

3.5.2. Oligonucleotide Probe Hybridization

The time required for duplex formation with oligonucleotide probes is minimized as they are single-stranded, of low complexity, and present in vast excess.

1. The filters are prehybridized individually in heat-sealed plastic bags for 30 min (with constant agitation) at 56°C in the 3*M* tetramethylammonium chloride hybridization buffer (*see* Note 4). Approximately 10 mL of solution/100 cm^2 of membrane should be used. Squeeze as much air as possible from the bag before sealing.
2. Remove the bag from the water bath. Open by cutting off one corner with scissors. Add the oligonucleotide probe (approx 1 pmol or 2×10^6 cpm/mL) directly to the prehybridization mix. Remove as much air as possible and reseal. (To avoid contaminating the water bath, the hybridization bag should be sealed inside a second, noncontaminated bag). Continue the hybridization at the same temperature for 1 h.
3. Remove the bag from the water bath, dry the outside and cut off one corner. Pour the hybridization solution into a beaker for disposal. Then cut the bag along the length of three sides and remove the filter, using a blunt forceps. Immediately submerge the filter in a tray containing 2× SSPE and 0.1% SDS. Wash for 10 min at room temperature with gentle shaking. Repeat.
4. Rinse the filter in 3*M* M_4NCl hybridization buffer minus carrier DNA and Denhardt's solution. Transfer the filter to a flat-bottomed lidded box and wash in this solution for 30–60 min at 60°C. (Final wash temperature suitable for a 20-base oligonucleotide probe, T_d 61–62°C.) The actual temperature of the wash solution itself should be monitored carefully (*see* Note 2).
5. Rinse the washed filter in 2× SSPE and wrap the damp filter in Saran Wrap™. **Important:** do not allow the filter to dry out at any stage if it is to be reused.
6. Cover the filter with a second piece of Saran Wrap™ and autoradiograph at –70°C (Kodak XAR film) using an intensifying screen for 1–12 h. Quantitative analysis of dot blots can be performed by densitometry of preflashed X-ray film.

3.5.3. Removing Probes from Nylon Filters

1. Immerse the membrane(s) in several hundred milliliters of 50% formamide, 5× SSPE, and 0.1% SDS. Incubate, with shaking, at 60°C for approx 15 min.

2. Rinse thoroughly with 2× SSPE at room temperature.
3. Wrap the damp membrane in Saran Wrap™ and autoradiograph to check that all the probe has been removed.
4. The membrane may now be dried or stored damp at 4°C until required.

4. Notes

1. "Spotty" background is usually the result of probe contamination with unincorporated [γ-^{32}P] ATP. "Blotchy" background is probably caused by poor washing. High background will be obtained if water from the outside of the hybridization bag is allowed to come in contact with the filter itself.
2. In M_4NCl, the T_d of a hybrid is dependent primarily on its length. However, very G+C rich oligonucleotides may be "sticky" i.e., have a T_d slightly higher than would normally be expected.
3. The length of oligonucleotides used as probes should be checked intermittently by polyacrylamide gel electrophoresis in case of degradation (and consequent signal loss).
4. SSPE (or SSC) may be used instead of M_4NCl for hybridization, in which case, the T_d depends on both the length and base composition of the probe.
5. Some workers purify radiolabeled oligonucleotides by precipitation with ethanol and ammonium acetate *(6)*.
6. This screening procedure for point mutations may be carried out on PCR amplified crude cell lysates *(2)*.
7. Where high SA are required, or probes of identical SA, polyacrylamide gel electrophoresis in the presence of 7*M* urea allows the separation of the "hot" from the "cold" oligonucleotide *(6)*.
8. It is possible to obtain probes with SA 5–10 times those achieved by the kinase reaction by "Primer Extension." This method uses the Klenow fragment of *E. coli* polymerase I and a short primer (usually an 8-mer), complementary to the 3'-end of the synthetic oligonucleotide, to incorporate [α-^{32}P]dNTPs in a template-directed manner *(6)*. The template and newly synthesized strands must then be separated by polyacrylamide gel electrophoresis under denaturing conditions.
9. Oligonucleotides can be used to probe genomic restriction digests by hybridization directly in dried agarose gels. However, extremely high SA probes are required, 20–30 µg DNA per slot for single copy genes and long autoradiographic exposures. This approach has now largely been superseded by the protocol outlined in this chapter.

10. Nonradioactive oligonucleotide probes have recently been introduced, e.g., probes covalently labeled with horseradish peroxidase, which allows detection with a simple colorimetric assay *(7)* (*see* Chapter 16).
11. A recent modification to the protocol outlined in this chapter has been to "reverse" the DNAs, i.e., attach the oligonucleotides to the nylon support and hybridize the amplified sample to the membrane *(8)*. This should be particularly valuable where the number of potential sequence variations exceeds the number of samples tested.

References

1. Wallace, R. B , Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T , and Itakura, K. (1979) Hybridization of synthetic oligodeoxyribonucleotides to ΦX174 DNA: The effect of single base pair mismatch *Nucleic Acids Res* **6**, 3543–3557.
2. Saiki, R K., Bugawan, T. L., Horn, G T , Mullis, K. B , and Erlich, H A. (1986) Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes *Nature* **324**, 163–166
3. Itakura, K., Rossi, J. J., and Wallace, R. B. (1984) Synthesis and use of synthetic oligonucleotides. *Ann Rev Biochem.* **53**, 323–356
4. Wood, W. I., Gitschier, J , Lasky, L. A., and Lawn, R M (1985) Base composition-independent hybridisation in tetramethylammonium chloride: A method for oligonucleotide screening of highly complex gene libraries *Proc Natl. Acad Sci USA* **82**, 1585–1588.
5. Jacobs, K. A , Rudersdorf, R., Neill, S. D., Dougherty, J. P , Brown, E L., and Fritsch, E. F. (1988) The thermal stability of oligonucleotide duplexes is sequence independent in tetralkylammonium salt solutions: Application to identifying recombinant DNA clones *Nucleic Acids Res.* **16**, 4637–4650
6. Sambrook, J , Fritsch, E F , and Maniatis, T (1989) *Molecular Cloning A Laboratory Manual*, 2nd Ed , Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
7. Saiki, R. K., Chang, C-A., Levenson, C. H., Warren, T. C., Boehm, C. D., Kazazian, H. H., and Erlich, H A. (1988) Diagnosis of sickle cell anaemia and β-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotides. *N Engl. J. Med* **319**, 537–541.
8. Saiki, R. K., Walsh, P. S., Levenson, C., and Erlich, H A (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes *Proc Natl. Acad. Sci USA* **86**, 6230–6234.

CHAPTER 8

Detection of Mutations by the Amplification Refractory Mutation System (ARMS)

John M. Old

1. Introduction

The amplification refractory mutation system (ARMS) is a simple and rapid method of detecting point mutations, restriction fragment length polymorphisms (RFLPs), and small nucleotide insertions or deletions. The method was first described by Newton et al. *(1)* for analyzing single DNA base differences in patients with α-1-antitrypsin deficiency and has since been applied to prenatal diagnosis and carrier detection of cystic fibrosis *(2)* and β-thalassemia *(3)*.

1.1. Principle of the ARMS Method

The technique is based on allele-specific priming of the polymerase chain reaction. To diagnose a specific mutation, two oligonucleotide primers are required that are identical in sequence except for the terminal 3' nucleotide. The normal primer has the 3' terminal nucleotide sequence complementary to the normal DNA sequence and the mutant primer has its 3' terminal nucleotide complementary to the mutant DNA sequence. Under the right conditions a primer will act as a template for DNA polymerase only when the terminal 3' nucleotide is perfectly matched to the target DNA sequence. Thus the normal primer when hybridized to the mutant genomic DNA and, conversely, the mutant primer when hybridized to the normal DNA, will not function properly as a template for DNA amplification and no amplified product is

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

observed at the end of the PCR process. To screen for a particular mutation, only the mutant ARMS primer is required (in combination with three others—a common primer and two control primers). However, for prenatal diagnosis of individuals homozygous for a particular mutation, both normal and mutant primers will be required.

The difficult part of setting up this technique is the design of the ARMS primers and testing them to establish that they are working properly. The technique itself is very simple, involving just a PCR assay and examination of the amplification products by agarose gel electrophoresis and ethidium bromide staining. The whole process takes only 4 h and thus disease-specific mutations can be quickly identified in DNA from individuals once a battery of mutation-specific ARMS primers have been developed *(4)*.

The technique requires two polymerase chain reactions involving four primers in the one reaction mixture. Two of the primers are control primers that amplify a segment of DNA some distance away from the site of the mutation under investigation so that they do not interfere with the amplification of the DNA fragment produced by the ARMS primer and common primer. The control fragment acts as an indicator that the PCR reaction was set up properly with all the right components and that the genomic DNA sample under investigation is of the right purity and concentration to permit amplification. When the control fragment is not observed in an ARMS analysis, the result obtained with the ARMS primer must be discounted and the analysis repeated. The other two primers in the reaction mixture are the ARMS primer, which is allele-specific to either the mutant DNA sequence or the normal DNA sequence as required, and a common primer, which matches the same sequence in both normal and mutant DNA. These two primers combine to produce the diagnostic ARMS DNA fragment in the amplification reaction.

Figure 1 illustrates the use of the ARMS technique to detect a point mutation (G→C) in the β-globin gene, which causes β-thalassemia. The nearest of the two control primers is situated approx 1000 base pairs distant from the site of the mutation. The ARMS primer to detect the mutation is complementary to the DNA sequence and has a G for its 3' terminal nucleotide in order to base-pair with the point mutation (C). When the ARMS primer sequence is complementary to the genomic sequence, the common primer has to be of the same sequence and lie 5' to the site of the mutation at a convenient distance for efficient amplification (300–500 bp is ideal). If the ARMS primer sequence is made identical to the genomic DNA sequence, then the common primer has to lie 3' to the mutation site and its sequence must be complementary to the genomic sequence.

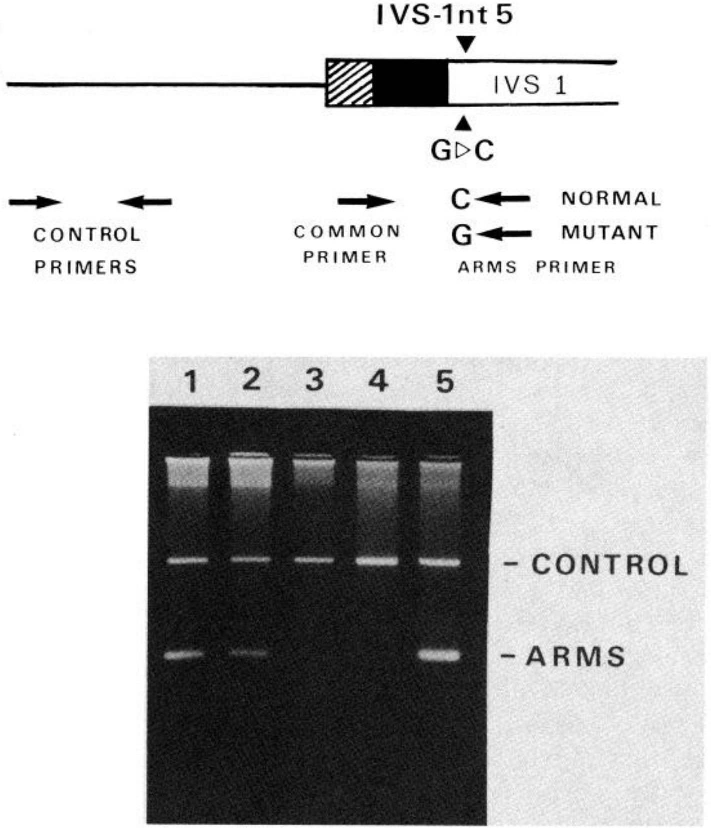


Fig. 1. Diagram of the 5' end of the β-globin gene showing the location of four primers required for an ARMS reaction. The position of a β-thalassemia point mutation (G to C) at nucleotide 5 in intervening sequence 1 (IVS1) is shown. The ARMS primers are complementary and reversed; the primer to detect the mutation has a G at its 3' terminus and the primer to detect the normal DNA sequence has a C. The stained agarose gel shows a prenatal diagnosis for the IVS1-nt5 mutation using the aforementioned primers. Track **1:** DNA from individual heterozygous for IVS1-nt5 (mother) with mutant ARMS primer. Result is positive. Track **2:** DNA from individual heterozygous for IVS1-nt5 (father) with mutant ARMS primer. Result is positive. Track **3:** DNA from normal individual with mutant primer. Result is negative. Track **4:** Fetal DNA with normal primer. Result is negative. Track **5:** Fetal DNA with mutant primer. Result is positive. The result from track 4 indicates that no normal DNA sequence is present at IVS1 nucleotide 5 and therefore the diagnosis is that the fetal DNA is homozygous for the IVS1-nt5 mutation.

1.2. Design of Primers

The success of the ARMS technique is dependent on the ARMS primers working specifically. After synthesis and purification it must be ensured that the primers do not allow amplification to proceed when they form a mismatched template. In many instances just a single 3'-terminal nucleotide mismatch will not stop amplification proceeding and therefore ARMS primers are usually designed with an additional deliberate mismatch of a single base at three or four nucleotides from the 3' end. In our experience the most suitable mismatch and its position is found by trial and error although there are some rules of guidance. According to Newton et al. *(1)* purine/purine and pyrimidine/pyrimidine mismatches have a greater effect than purine/pyrimidine mismatches. In particular, a C/C mismatch has a greater effect than C/T, then C/A, and finally G/T has the weakest destabilizing effect.

Primers designed to detect a mutation can be tested with normal genomic DNA samples to determine whether they are working specifically. Occasionally, a primer may not be found to work satisfactorily even after several attempts at introducing the additional mismatched base. In such cases, with primers for β-globin gene mutations, we have found that reversing the sequence of the primer in order to amplify in the opposite direction has solved the problem *(3)*. Primers designed to detect the normal DNA sequence at the site of a mutation must be tested using DNA from an individual homozygous for that particular mutation.

Primers must also be shown to amplify efficiently with the correct target DNA. If a series of primers are to be used at the same time, for example for screening a DNA sample for one of many different mutations, it is convenient to ensure that the primers will all work efficiently at the same set of reaction conditions and annealing temperature. The latter is dependent on the length of the primers. Initially we tried using 24-nucleotide primers for the ARMS technique with modest success; however, increasing the length to 30 nucleotides (as recommended in ref. *1*) permitted a 65°C annealing temperature to be used. This prevented any nonspecific bands and also seemed to improve the specificity of the ARMS primers. It also allows the three-step amplification cycle to be reduced to just two steps of 93°C and 65°C if required for convenience.

Finally, the design of the common primer and two control primers should not be neglected. The control amplified fragment should be of sufficient size difference to the ARMS fragment to allow easy separation by agarose gel electrophoresis. It is also very important to ensure that there are no other known mutations or polymorphic DNA sequence changes located in the DNA sequences chosen for any of the four primers in the ARMS assay. Any such

occurrence is liable to reduce the effectiveness of the primer and may lead to an erroneous result.

2. Materials

1. Prepare a stock of 10× Cetus buffer by adding together 0.5 mL of 1*M* Tris-HCl (pH 8.3), 1.25 mL 2*M* KCl, 75 µL 1*M* $MgCl_2$, 5 mg gelatin and 3.275 mL of distilled water to make a total volume of 5 mL. The buffer should be heated to 37°C until the gelatin dissolves and can then be kept frozen.
2. Prepare a stock of deoxynucleotide mixture (1.25 m*M* each dNTP) by adding together 50 µL of 100 m*M* dATP, 50 µL of 100 m*M* dCTP, 50 µL of 100 m*M* dGTP, 50 µL of 100 m*M* dTTP, and 3.8 mL of distilled water. Deoxynucleotide solutions must be of neutral pH and can either be prepared oneself by dissolving each from powder, carefully neutralizing the solution with NaOH until pH 7, measuring the absorbance at 260 nm (*see* Chapter 7), and diluting to make a final concentration of 100 m*M*, or one can buy ready made dNTP solutions already neutralized and diluted to 100 m*M* (e.g., Boehringer Mannheim, East Sussex, UK) at greater cost but much greater convenience.
3. Prepare a reaction mixture stock (4 mL) by adding together 0.5 mL 10× Cetus buffer, 0.8 mL 1.25 m*M* dNTP mixture, and 2.7 mL distilled water.
4. Dilute primer stock solutions to make working solutions at a concentration of 1 OD U/mL.
5. AmpliTaq® DNA polymerase. This is supplied by I.L.S. Limited (London; a subsidiary company of Perkin-Elmer Limited).

3. Methods

3.1. PCR Conditions

The standard PCR conditions recommended by Cetus for their AmpliTaq® enzyme can be used. These are: 10 m*M* Tris-HCl, pH 8.3, 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 0.01% gelatin, 0.2 µ*M* each primer, and 200 µ*M* each of dATP, dCTP, dGTP, and dTTP. The reaction volume can be scaled down from the recommended 100 µL to just 25 µL, which then requires only 0.5 units of AmpliTaq® enzyme and 0.5 µg of genomic DNA to complete the reaction.

Set up the reaction mixture as follows (*see* Note 1):

1. Pipet 20 µL of the reaction mixture (*see* Materials 3) into a 0.5-mL Eppendorf tube.

2. Add 1 µL of each control primer, 1 µL of the common primer, and 1 µL of the appropriate ARMS primer.
3. Add 0.5 units of AmpliTaq® enzyme. This is supplied at a concentration of 5 U/µL and therefore only 0.1 µL is required. This is difficult to estimate and therefore it is best to make a dilute solution of enzyme at a concentration of 0.5 U/µL using distilled water.

   If the same four primers are being used in every tube (for example when screening a number of DNA samples for one mutation), then all four primers and enzyme can be mixed together (1 µL each primer and 0.1 µL U enzyme per tube) and then 4.1 µL of primer/enzyme mixture can be added to each tube.
4. Add 1 µL of DNA solution (at approx 0.5 mg/mL).
5. Add 25 µL of light paraffin oil.
6. Place the tubes in a PCR machine and program for 25 cycles at:
   a. 93°C for 1 min
   b. 65°C for 1 min
   c. 72°C for 1.5 min

   with a final extension period of 3 min.

   The annealing temperature of 65°C is appropriate for primers of 30 nucleotides in length (*see* Design of Primers). No initial denaturation step is necessary before the program of 25 cycles.

3.2. Analysis of PCR Product

1. Remove the tubes from the PCR machine and pipet 5 µL of ficoll/bromophenol blue dye (15%/0.05%) into each tube.
2. Vortex the mixture to incorporate the blue dye and spin for 5 s in a microfuge.
3. Remove a 20-µL aliquot of the blue aqueous mixture underneath the paraffin oil layer, and load into the well of a submerged agarose minigel. For most purposes a 3% agarose gel is used made up of 1.5% agarose and 1.5% Nusieve agarose (FMC BioProducts, Kent, UK, marketed through Flowgen).
4. After electrophoresis at 100 V for approx 45 min in standard Tris-Acetate buffer, the gel is stained in ethidium bromide solution and photographed on a UV transilluminator.

The stained agarose gel should look like the one shown in Fig. 1. An amplified fragment from the control primers is present in each of the five tracks indicating that the DNA samples and reaction conditions permitted amplification to take place (*see* Note 2). A positive result showing ampli-

fication by the ARMS and a common primer is observed in tracks 1, 3, and 5. The DNA samples and type of ARMS primers used are explained in the legend.

4. Notes

1. Like most PCR techniques the ARMS method is very sensitive, and therefore the possibility of error through contamination of the reaction mixture with either amplified or unamplified genomic DNA is always present. All the usual precautions should be taken when setting up the reaction mixtures. In particular, a dedicated set of automatic pipets is required which is never used for handling amplified products. For prenatal diagnosis an additional problem is the possibility of maternal DNA contamination in the fetal sample. All chorionic villus samples must be sorted free of any maternal decidua and blood clots by microscopic dissection before making DNA. Likewise amniotic fluid samples must never be bloodstained, otherwise there is a risk of maternal DNA being present. For prenatal diagnosis of β-thalassemia we always use RFLP linkage analysis in addition to the direct detection of mutations by the ARMS technique whenever possible in order to provide a confirmatory result because of this problem.
2. As with other PCR techniques, a degree of experience is required to interpret the results. The amplified fragment produced by the ARMS and common primer should be more or less equal in quantity to the amplified control fragment. Occasionally an amplified ARMS fragment much fainter than the control band is observed. This is usually an artifact from a slight loss of specificity of the primer, probably caused by a variation of the reaction conditions. This is easily demonstrated by always running both positive and negative control DNA samples. However, it can cause problems when the control amplified fragment is itself very faint. This occurs when a DNA sample is very dilute and it is best to concentrate the sample by reprecipitation and repeat the assay, rather than simply repeating the reaction for a larger number of cycles and increasing the risk of a possible DNA contamination problem. Finally, with some DNA samples the control fragment is completely absent and no interpretation of the ARMS result should be made. This problem can be overcome by cleaning up the DNA sample by phenol/extraction. However, if the DNA sample is too small and precious for such treatment, the addition of spermidine to the reaction mixture at a final concentration of 1 m*M* usually allows amplification to proceed.

References

1 Newton, C. R , Graham, A., Heptinstall, L. E., Powell, S J , Summers, C., Kalsheker, N., Smith, J. C., and Markham, A. F (1989a) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17,** 2503–2516.

2. Newton, C. R , Heptinstall, L. E., Summers, C., Super, M , Schwartz, M , Anwar, R., Graham, A., Smith, J. C , and Markham, A. F. (1989b) Amplification refractory mutation system for prenatal diagnosis and carrier assessment in cystic fibrosis. *Lancet* ii, 1481–1483.

3. Old, J. M., Varawalla, N. Y, and Weatherall, D. J. (1990) Rapid detection and prenatal diagnosis of β-thalassaemia: Studies in Indian and Cypriot populations in the UK. *Lancet* ii, 834–837.

4. Varawalla, N. Y., Old, J M , Sarkar, R., Venkatesan, R., and Weatherall, D J. (1991) The spectrum of β-thalassaemia mutations on the Indian subcontinent: The basis for prenatal diagnosis. *Br. J. Haematol.,* in press.

CHAPTER 9

Automated Gene Detection Using the Oligonucleotide Ligation Assay

Ulf Landegren

1. Introduction

Routine DNA analysis is assuming an increasing importance in clinical medicine as well as in basic research. The expanding field of DNA diagnostics will take advantage of the growing understanding of the molecular basis of human disease. Similarly, extensive studies of the structure of large genomes require efficient analytical techniques. In this chapter a strategy will be described that permits rapid distinction between alternate DNA sequences under standardized conditions. Applications of the technique includes screening for mutant gene sequences and genetic linkage analysis. The analysis may be performed manually or by a laboratory workstation resulting in a generally applicable automated gene analysis technique.

Large scale molecular genetic analyses require automated procedures *(1)*. Efficient and economical analytical procedures will render, as practical, routine prediction of genetic disease or increased susceptibility to disease, genetic linkage studies, pathological classification in malignancy, and the forensic identification of individuals. Similar techniques will also be critical for the identification of infectious organisms and in veterinary medicine and plant breeding. We have taken advantage of the standardized formats of two recently developed gene analysis techniques, the polymerase chain reaction (PCR) *(2,3)* and the oligonucleotide ligation assay (OLA) *(4)*, by applying these together in a scheme for the analysis of DNA sequence variants *(5)* (Fig. 1).

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by C Mathew Copyright © 1991 The Humana Press Inc , Clifton, NJ

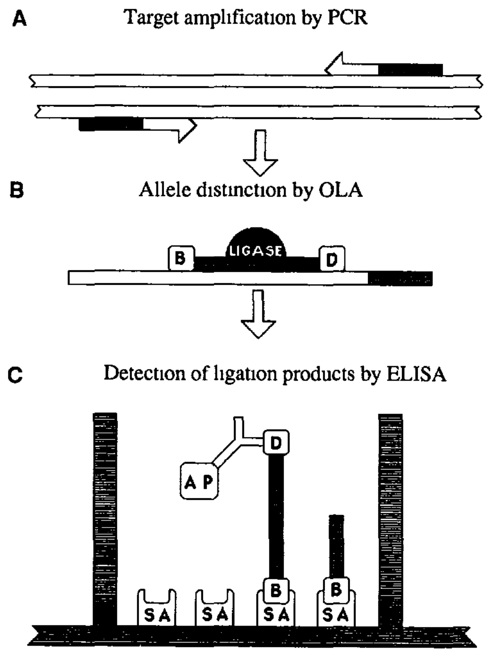


Fig. 1. A schematic outline of the strategy to automatically analyze gene sequence variants. The gray rectangles represent synthetic oligonucleotides. **A:** The oligonucleotides prime the synthesis of new strands in a PCR. **B:** Allelic variants of the newly synthesized strands, are distinguished in the OLA reaction. This step investigates the ability of a DNA ligase to join two oligonucleotides, hybridizing in juxtaposition on a target strand. The 5' and 3' oligonucleotides are modified by the addition of biotin (B), and digoxigenin (D) residues, respectively. **C:** Ligation products are detected using an enzyme-linked immunosorbent assay (ELISA). Biotinylated oligonucleotides are immobilized in microtiter wells, coated with streptavidin (SA). Digoxigenin-modified oligonucleotides, immobilized as a consequence of having been ligated to biotin-labeled ones, are detected using antibodies to digoxigenin, coupled to alkaline phosphatase (AP).

The PCR technique permits the exponential accumulation of DNA segments defined by two flanking primer sequences. With this technique even very limited tissue samples may be analyzed and DNA sequences may be detected in an enormous excess of irrelevant DNA. In the OLA technique the enzymatic joining of two synthetic oligonucleotide probes, hybridizing to a target strand in immediate juxtaposition, depends on the correct base pairing of the junction region of the two oligonucleotides. Thus, sequence variants may be rapidly distinguished under standardized conditions. The successful ligation can be visualized by retrieving one of the two oligonucleotides via an

added biotin residue and assaying for the coimmobilization of the other oligonucleotide, labeled with a detectable group. We use as a detectable group a hapten (digoxigenin) that can be monitored by enzyme-linked immune detection. The combination of target amplification by PCR with sequence distinction using OLA results in a sensitive, robust automatable assay for the presence of DNA sequence variants *(5)*.

2. Materials

2.1. General

1. All procedures can be performed manually but for increased efficiency and reproducibility a laboratory workstation (Biomek 1000, Beckman Instruments, Palo Alto, CA) may be programmed to execute the sample handling steps. This instrument is used with multipipet tools and with a multibulk tool to perform the reagent additions and washing steps of the assay. The pipetting tools used to assemble the components for PCR should not be used for the subsequent processing of the reactions in order to avoid contamination.
2. DNA samples and reagents are kept in 96-minitube cassettes (titertubes and racks, Bio-Rad, Richmond, CA).
3. All reactions during the assay are performed in flexible microtiter plates.

2.2. Amplification of a Target DNA Fragment

1. Genomic DNA may be purified by standard proteinase K digestion and phenol extraction *(6)* and stored at 2 ng/µL in water containing 0.1% Triton X-100 (TX100). Alternatively, DNA samples may be obtained by scraping cells from the cheek mucosa of human subjects with a toothpick. The cells are added to minitubes containing 50 µL of water with 0.1% TX100, overlayered with 80 µL light weight mineral oil and boiled for 5 min. The samples are stored at –20°C until use.
2. Twice concentrated PCR mixes typically contain 20 m*M* Tris-HCl, pH 8.3, 100 m*M* KCl, 3 m*M* $MgCl_2$, 12.5 ng/µL bovine serum albumin, the four deoxynucleotides at 400 µ*M* each, 1 µ*M* of each of the two PCR primers, 0.1% TX100, and *Taq* polymerase, 10 U/mL added fresh to the mixes shortly before use. For individual PCR reactions the $MgCl_2$ concentration may have to be varied in steps of 0.5 m*M* to optimize the amplification reaction.
3. The temperature is controlled by a thermal cycler capable of amplifying samples in microtiter wells.

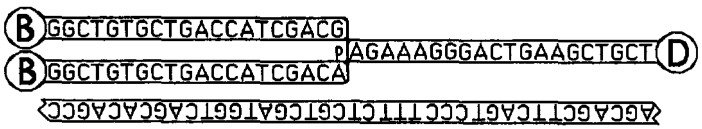


Fig. 2. Nucleotide sequence of oligonucleotides employed for the distinction of the normal and the Z mutation of the α-1-antitrypsin gene (in 5' to 3' orientation). The two 5' oliogonucleotides are modified by the addition of a biotin (B) residue in a 5' position. They differ from each other in the 3'-most position reflecting the difference between the normal and the Z variant of the α-1-antitrypsin gene. Whichever of these two oligonucleotides that is correctly matched to the target sequence can then ligate to the 3' oligonucleotide, bearing a 5' phosphate group (p) and with digoxigenin residues (D) added to its 3' end. The target sequence of the normal allele of the α-1-antitrypsin gene is also shown.

2.3. Oligonucleotide Ligation Analysis

1. Solutions for denaturation and for the subsequent restitution of pH: 0.3*M* NaOH containing 0.1% TX100 and 0.3*M* HCl with 0.1% TX100.
2. Three 20-mer oligonucleotides are used for the distinction between two alternative DNA sequences (*see* Fig. 2). In each ligation assay the presence of one of the two sequence variants is investigated. The ligation assay measures the ability of one allele-specific oligonucleotide to ligate to another oligonucleotide hybridizing to the target strand immediately downstream of the first oligonucleotide. The allele-specific oligonucleotides are modified in the 5' position by the addition of a biotin group. Most conveniently the biotin group may be added during the last step of the oligonucleotide synthesis (e.g., biotin dX, Midland Certified Reagent Company, Midland, TX, as described by the manufacturer). The oligonucleotide hybridizing immediately 3' of either of the two allele-specific oligonucleotides is modified by the addition of a 5' phosphate group, required for the ligation reaction, and by introducing the hapten digoxigenin in a 3' position. The 5' phosphate group may be added enzymatically or during synthesis using the "5'-Phosphate-On" reagent (Clonetech, Palo Alto, CA) as described by the manufacturer. For the addition of the digoxigenin group 500 pmol of phosphorylated oligonucleotide is incubated in 100 µL of 100 m*M* potassium cacodylate, pH 6.8, 2 m*M* $CoCl_2$, 200 µ*M* dithiothreitol, 2.5 µL dUTP-digoxigenin (Boehringer-Mannheim), and 70 U of terminal deoxynucleotidyl transferase for 1 h at 37°C. The oligonucleotide may be used without purification. Typically between one and three dUTP-digoxigenin residues are added per oligonucleotide molecule.

3. The twice-concentrated ligation mixture contains one of the allele-specific oligonucleotides and the oligonucleotide hybridizing immediately 3' of the allele-specific oligonucleotides (Fig. 2). The oligonucleotide pair is present at 20 n*M* each in 100 m*M* Tris-HCl pH 7.5, 265 m*M* NaCl, 20 m*M* $MgCl_2$, 2 m*M* spermidine, 2 m*M* adenosine triphosphate, 20 m*M* dithiothreitol, 20 ng/µL bovine serum albumin, and 2.5 Weiss units T4 DNA ligase per mL.

2.4. Enzyme-Linked Immune Detection of Ligation Products

1. Wells of microtiter plates are coated for 2 h at 37°C with 60 µL streptavidin (e.g., Sigma, St. Louis, MO), 100 µg/mL in phosphate-buffered saline. The wells are blocked for at least 30 min at room temperature with 200 µL/well of blocking buffer; 0.5% fat-free dry milk, 100 µg/mL of denatured salmon sperm DNA, and 0.02% (w/v) NaN_3 in buffer A. Buffer A consists of 100 m*M* Tris-HCl, pH 7.5, 150 m*M* NaCl, 0.05% Tween 20. The microtiter plates may be stored with blocking solution at +4°C.
2. Denaturing wash buffer: 0.01*M* NaOH and 0.05% Tween 20.
3. Before use, alkaline phosphatase-conjugated anti-digoxigenin antibodies (Boehringer-Mannheim) are diluted 1:1000 in buffer A.
4. The substrate solution for alkaline phoshatase is prepared fresh by dissolving a 5 mg tablet of paranitrophenylphosphate (e.g., Sigma) in 100 m*M* diethanolamine, 0.5 m*M* $MgCl_2$, pH 9.5.
5. The results of the enzyme-linked detection reaction are determined using a microtiterplate-reading spectrophotometer.
6. A personal computer is used to process the absorbance data.

3. Methods

3.1. Amplification of Target DNA

1. Add *Taq* polymerase shortly before use to the premixed solution for PCR. Distribute 5 µL aliquots of 2× PCR mixes in individual wells of a microtiter plate.
2. Add 5 µL of genomic DNA samples to each well and overlay the reactions with 70 µL of mineral oil (*see* Note 1).
3. Subject the reaction wells to temperature cycling, typically denaturing the DNA at 93°C for 30 s, annealing the oligonucleotides at 55°C for 45 s, and enzymatically extending the primers at 72°C for 35 cpm. Depending on the particular oligonucleotide primers used the annealing temperature may have to be modified.

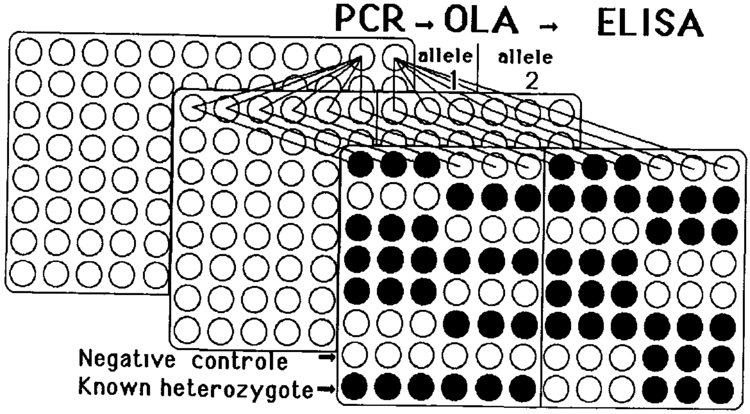


Fig. 3. Description of the sequential transfer of reactions in the present scheme for the analysis of DNA sequence variants. A PCR reaction established from one DNA sample is distributed to two triplicate sets of wells for the OLA-based identification of each of two allelic variants of the amplified sequence. Ligation products from each OLA reaction are then transferred and analyzed by ELISA. Each ELISA microtiterplate investigates 16 DNA samples for the presence of two different alleles.

3.2. Oligonucleotide Ligation Assay

1. Add 45 µL of 0.3*M* NaOH, 0.1% TX100 to the amplified DNA samples to denature the PCR products.
2. Add T4 DNA ligase fresh to OLA reaction mixes. Distribute 10 µL aliquots of OLA mixes designed to detect each of two allelic forms of the amplified DNA sequence to two sets of triplicate microtiter wells (Fig. 3).
3. Neutralize the amplified samples by the addition of 45 µL of 0.3*M* HCl, 0.1% TX100.
4. Immediately distribute six 10-µL portions of the denatured PCR samples to six wells containing the OLA reagents.
5. Incubate the ligation reactions for 15 min at 37°C.
6. Terminate the reactions by the addition of 10 µL 0.3*M* NaOH, 0.1% TX100 to each well.

3.3. Enzyme-Linked Immune Detection of Ligation Products

1. Wash a streptavidin-coated microtiter plate twice with buffer A to remove any free streptavidin.
2. Neutralize the ligation reactions by the addition of 10 µL 0.3*M* HCl, 0.1% TX100.
3. Transfer these reactions to a streptavidin-coated microtiter plate.

4. After 15 min at 37°C, wash the plate once with buffer A and then twice with 0.01*M* NaOH to dissociate oligonucleotide pairs that are joined by hybridization to the same target strand without having been ligated, and once again with buffer A.
5. Add 30 µL of antidigoxigenin antibodies in blocking buffer to each ELISA reaction well and incubate for 30 min at 37°C. Wash the wells six times in buffer A.
6. Add 50 µL of substrate solution per well and incubate at room temperature untill the spectrophotometric absorbance at 405 nm approaches an optical density of 2.0 (*see* Notes 3 and 6).
7. Enter the spectrophotometric reading in a personal computer. The data may be used to calculate and store the mean and standard deviation for each of two analyzed alleles of a DNA sample. Calculate the ratio of the means of the absorbance scored for the two alleles to determine if the individual is homozygous for either allele or if the individual is a heterozygote.

An example of an analysis using the PCR-OLA-ELISA (POE) procedure is presented in Fig. 4. DNA samples derived from individuals with α-1-antitrypsin gene defects (kindly provided by Dr. Fielding Hejtmancik) were investigated. The spectrophotometric signal was recalculated to indicate if an individual had two copies of the normal (filled) or the mutant (open) allele or if the sample was derived from a heterozygote with one copy of each allele.

4. Notes

1. Positive and negative controls should be included in each set of samples investigated. In general, contamination represents a serious concern in procedures involving PCR. We have not, however, noted this as a problem in the present automated analysis.
2. The OLA reaction appears capable of distinguishing target sequences differing by any point mutation under a standard set of conditions *(4)*. These conditions may have to be adjusted by the individual experimenter by varying the NaCl concentration (typically 200 m*M*) or the amount of ligase used during OLA. The inclusion of formamide during the ligation assay may also serve to enhance the distinction *(5)*.
3. If the assay produces a poor signal, each step should be analyzed separately. The amplification of target segments may be monitored by agarose gel electrophoresis.
4. The ligation of the oligonucleotides in the OLA step may be investigated, independent of the ELISA step, by labeling one of the two oligonucleo-

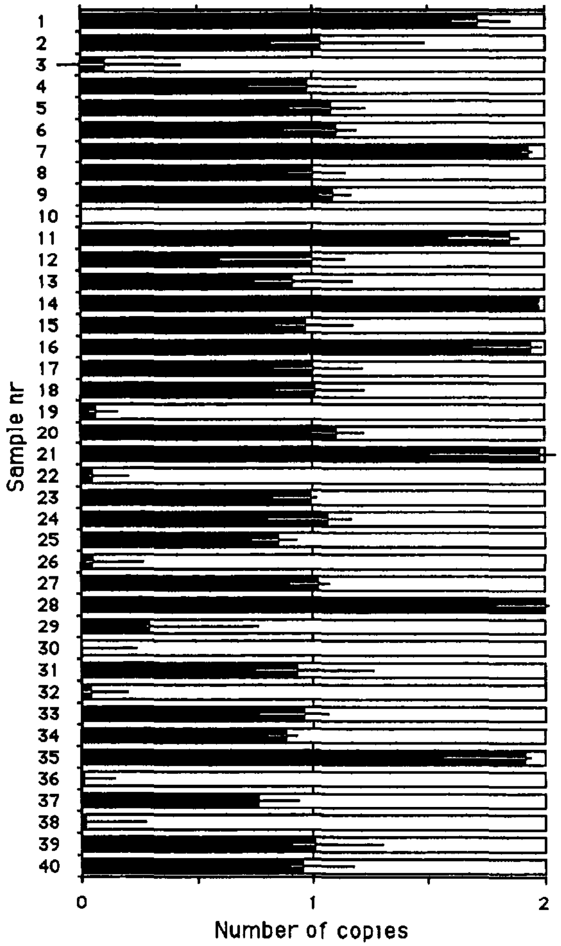


Fig. 4. Processed data from an analysis of the α-1-antitrypsin genotype in 40 individuals. The results have been computed to reveal if the individuals have no, one, or two copies of the normal (filled bars) or the mutant gene (open bars). The means of the triplicate determinations are given with standard errors. Standard errors for the determination of the normal allele extend into the filled bars and conversely. The absorbance data have been corrected so that the signal from the two allelic variants add up to the expected two gene copies per cell.

tides with a radioactive phosphate group in the 5' position using the enzyme polynucleotide kinase and γ-^{32}P-ATP. The ligation products are analyzed by electrophoresis in a 15% denaturing polyacrylamide gel and autoradiography *(6)*.

5. In order to characterize the enzyme-linked detection reaction it is convenient to label a 5' biotinylated oligonucleotide with dUTP-digoxigenin in the 3' position as described (Section 2.3.2.) to circumvent the requirement for the amplification and ligation steps of the assay. If the binding capacity of the streptavidin-coated microtiter wells appear insufficient, a higher loading may be achieved by first coating the wells with biotinylated bovine serum albumin before washing and adding streptavidin.
6. A stronger spectrophotometric signal may be obtained by using another substrate for the alkaline phosphatase (e.g., BRL-ELISA detection system *[5]*).
7. The present automated gene detection strategy permits processing 1200 ligation reactions a day *(5)*. Further developments that will serve to increase the efficiency of the analysis include differential labels for reagents specific for the two allelic variants, permitting the analysis of two allelic variants in a single assay well.

References

1. Landegren, U., Kaiser, R, Caskey, C. T., and Hood, L. (1988) DNA diagnostics—molecular techniques and automation *Science* **242**, 229–237.
2. Saiki, R., Scharf, S, Faloona, F., Mullis, K., Horn, G., Erlich, H, and Arnheim, N. (1985) Enzymatic amplification of ß-globin sequences and restriction site analysis for diagnosis of sickle cell anemia *Science* **230**, 1350–1354
3. Saiki, R, Gelfand, R., Stoffel, S, Scharf, S., Higuchi, R, Horn, G, Mullis, K., and Erlich, H (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.
4. Landegren, U., Kaiser, R, Sanders, J., and Hood, L. (1988) A ligase-mediated gene detection technique. *Science* **241**, 1077–1080.
5. Nickerson, D. A., Kaiser, R, Lappin, S., Stewart, J, Hood, L., and Landegren, U. (1990) Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay *Proc Natl. Acad Sci, USA* **87**, 8923–8927
6. Maniatis, T., Fritsch, E, and Sambrook, J (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

CHAPTER 10

Detection of Point Mutations by Denaturing-Gradient Gel Electrophoresis

Stephen R. Dlouhy, Patricia Wheeler, James A. Trofatter, Peter J. Stambrook, and Jay A. Tischfield

1. Introduction

Denaturing-gradient gel electrophoresis (DGGE) detects DNA sequence differences. Thus, it can be used to screen for point mutations or other types of mutation prior to DNA sequence analysis. The technique, first described by Fischer and Lerman *(1)*, entails electrophoresis of DNA fragments at high temperature (ca. 60°C) in an acrylamide gel that contains a gradient of denaturant (formamide and urea). As a DNA fragment migrates in the gel, it encounters increasing concentrations of denaturant and at some point, it will become partially or totally single-stranded (melted, denatured). The position at which the DNA melts is determined by its nucleotide sequence and composition. Partial or complete denaturation causes a marked decrease in the electrophoretic mobility of the DNA and results in bands that are usually quite sharp. Two DNA fragments of the same size but of different sequence frequently will denature at different points within the gradient, and will therefore be separable by DGGE. In general, A-T-rich sequences denature at lower denaturant concentrations than G-C rich sequences. The method is applicable for fragments that are about 50–1000 bp in length *(2,3)*.

From *Methods in Molecular Biology, Vol. 9 Protocols in Human Molecular Genetics*
Edited by· C Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

DGGE is useful as a rapid screen for mutations in a gene, although it does not establish the precise position or type of mutation(s) involved. Rather, it provides a comparative analysis between samples. The technique is versatile because sequence changes can be detected even if they are not at restriction endonuclease recognition sites. However, because DNA fragments denature in distinct segments or "domains," only those sequence differences that reside within the early melting domains of a particular fragment can be detected by DGGE *(2–4)*. In addition, some base substitutions in native duplex DNA do not result in differences in denaturation even if they reside within an early melting domain *(5–7)*. Thus, although the separation of two fragments by DGGE can be indicative of a sequence difference between them, failure to observe separation does not exclude such differences. To overcome this difficulty, the efficiency of detection of mutations/sequence differences by DGGE can be increased by (1) the production of heteroduplexes, (2) digestion of the DNA with different restriction enzymes to alter melting domains, or (3) the attachment of terminal GC-rich regions known as GC clamps (*see* Methods). DGGE has been applied to the detection of mutations in cloned and genomic DNAs, including DNA amplified by the polymerase chain reaction (PCR) (*8–14*).

2. Materials

2.1. Apparatus

1. Overview of the apparatus: It is important that the gel be uniformly heated. This is accomplished by submerging it in heated, well-mixed buffer. Furthermore, the electrophoresis buffer must be recirculated in order to maintain constant pH. Several commercially available electrophoresis systems are potentially suitable for DGGE, including units from Hoefer (SE 600, SE 620, or SE 660), Bio-Rad (Protean II), or Green Mountain Lab Supply (Waltham, MA). Alternatively, one can construct a suitable apparatus *(2)*. For brevity and simplicity, only the use of a commercial system will be described.

   A schematic of a DGGE apparatus, following the basic principles of a Hoefer SE 660 unit, is shown in Fig. 1. The major components of the system are:

   a. An electrophoresis chamber, including the gel and a heat exchanger.
   b. A pump for recirculating the electrophoresis buffer.
   c. An external, circulating, heated water bath.
   d. An electrophoresis power supply capable of generating up to 250 V and up to 150 mA.

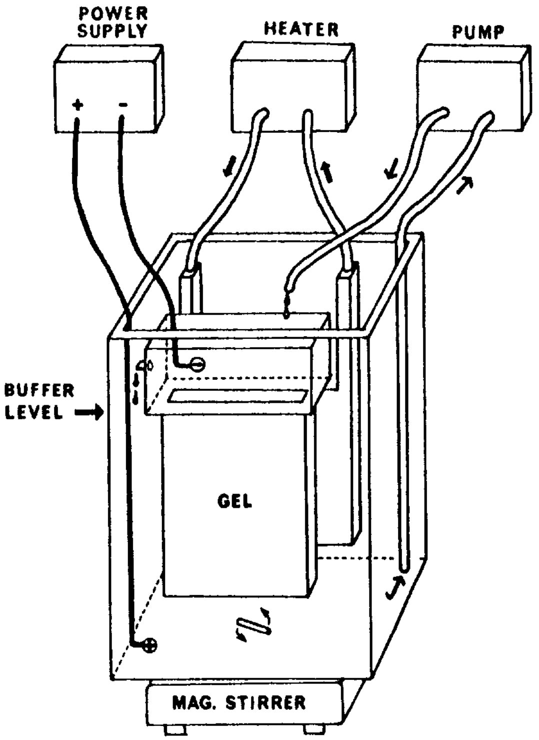


Fig. 1. Schematic representation of the major components of an apparatus used for denaturing gradient gel electrophoresis *(see text for details)*. Although only a single gel is shown, with some electrophoresis systems it is possible to run multiple gels simultaneously. For simplicity, some aspects of the apparatus are not shown, such as the cover, insulating Styrofoam (optional), and supports for the heat exchanger and gel assembly The buffer level for the lower (outer) chamber is indicated.

To reduce heat loss from the apparatus, the chamber can be insulated with Styrofoam sheets (1–2 cm thick) on the outside of the buffer chamber.

2. Electrophoresis chamber: Lower and upper buffer chambers, heat exchanger, and gel preparation materials (glass plates, spacers, combs, and pouring frame), electrodes, and so on, are part of the commercial electrophoresis systems, and are assembled according to the supplier's instructions.
3. Buffer recirculation pump: It is best to use a peristaltic or equivalent pump with the capability of varying the flow rate. Recirculation of buffer at 200 mL/min works well for the Hoefer SE 660.

4. Circulating, heated water bath (Heater): When using an external, circulating, heated water bath, heat loss occurs between the water bath and gel chamber. Use of a Neslab RT3-210D circulating water bath (flow rate ca. 3 L/min) in conjunction with an insulated Hoefer SE 660 results in about a 3–4°C drop in temperature between the heated bath and the interior of the electrophoresis chamber. Any comparable apparatus should suffice. Alternatively, for a homemade apparatus, an aquarium style heater can be used directly in the chamber.
5. Miscellaneous materials: Tubing, clamps, and connections for buffer recirculation and for external circulating, heated water bath; gradient maker (chamber vol ca. 10–30 mL); magnetic stir plate with a surface area large enough to stably support the electrophoresis chamber; UV light box; thermometer for inside the electrophoresis chamber; microsyringe (e.g., Hamilton); micropipet (e.g., Pipetman); blotting membrane (e.g., Nytran); and an electroblot chamber, such as a Bio-Rad Trans-Blot. Tubing for connecting the circulating heater should be heat-resistant (e.g., autoclavable) since it will expand with prolonged use.

2.2. Solutions

1. 40% Acrylamide stock solution (37.5:1 acrylamide:bisacrylamide): Dissolve 194.8 g electrophoresis grade acrylamide and 5.2 g bisacrylamide in distilled water and bring vol to 500 mL. Keep in a brown glass bottle at 4°C. CAUTION: Acrylamide is a neurotoxin, use gloves and surgical mask when handling powder, and gloves when handling solutions.
2. 50x stock TAE gel buffer: 2*M* Tris, 1*M* sodium acetate, 50 m*M* EDTA, pH 7.4. The pH is adjusted with glacial acetic acid.
3. 100% Denaturant stock solution: 6.5% Acrylamide, 40% formamide, and 7*M* urea in 1x TAE buffer. Mix 81.25 mL of 40% acrylamide stock, 10 mL of 50x TAE buffer, 210 g urea, and 200 mL deionized formamide. Fluka formamide (47670) can be used directly without additional deionization. Adjust the final vol to 500 mL.
4. Denaturant-free stock solution: 6.5% Acrylamide in TAE buffer. Mix 81.25 mL of 40% acrylamide stock, 10 mL of 50x TAE buffer, and 468.75 mL distilled water.
5. Ammonium persulfate (APER, $(NH_4)_2S_2O_8$) (10%): Bring 1 g ammonium persulfate to 10 mL with distilled water. Make fresh. CAUTION: APER is a strong oxidant.
6. TEMED (*N,N,N′,N′*-tetramethylethylenediamine). CAUTION: TEMED is flammable and causes irritation.

7. Sample loading solution (5x): 80% glycerol, 10 m*M* Tris-HCl, pH 7.4, 25 m*M* EDTA, 0.25% bromphenol blue.

3. Methods

3.1. Preliminary Considerations

There are variations in the design and preparation of denaturing gradient gels, and in the treatment of samples prior to electrophoresis. Each has its advantages and shortcomings relevant to specific applications.

1. Orientation of denaturing gradient: A DGGE gel can be oriented so that the denaturant gradient is either parallel with or perpendicular to the direction of electrophoresis. The parallel method is more useful for side by side comparisons of multiple samples, and will be described in detail. Perpendicular gradient gels are discussed in Section 4.1.
2 Denaturation point and gradient concentration range: In order to make gradient gels that are optimal for detection of mutation in a particular fragment, it is necessary to determine the approximate denaturant concentration required to cause partial melting/denaturing of the fragment. This concentration can be established empirically with a parallel gradient gel(s) *(see below)* or with a perpendicular gradient gel (*see* Section 4.1). Alternatively, one can use a computer algorithm to predict melting behavior *(3,4)*.

   To estimate the denaturation point with a parallel gradient gel, use moderate voltage (60–90 V); a broad concentration range of denaturant (e.g., 10–90%); electrophoresis of long duration (18–24 h); and multiple, temporally staggered loads of the same sample (e.g., load at $t = 0$, 5, and 10 h). Alternatively, the same sample can be loaded onto two different gradients (e.g., 10–60% and 50–90%) and electrophoresis conducted at higher voltage (130–160 V) for shorter periods (6–12 h). The appearance and position of bands at the end of the electrophoresis provides data concerning the denaturation point of the fragment. Fragments will be significantly retarded at the concentration of denaturant that is sufficient to cause melting of early melting domains. Also, when denatured, the DNA bands do not trail at the sides of the channel (i.e., "smile") and are usually quite sharply focused (e.g., Fig. 2A). The denaturing concentration for any fragment is estimated by determining the position in the gradient at which the fragment denatured. For example, if a fragment denatures approximately halfway down a 20–60% gel, it is denaturing at a concentration of about 40% denaturant.

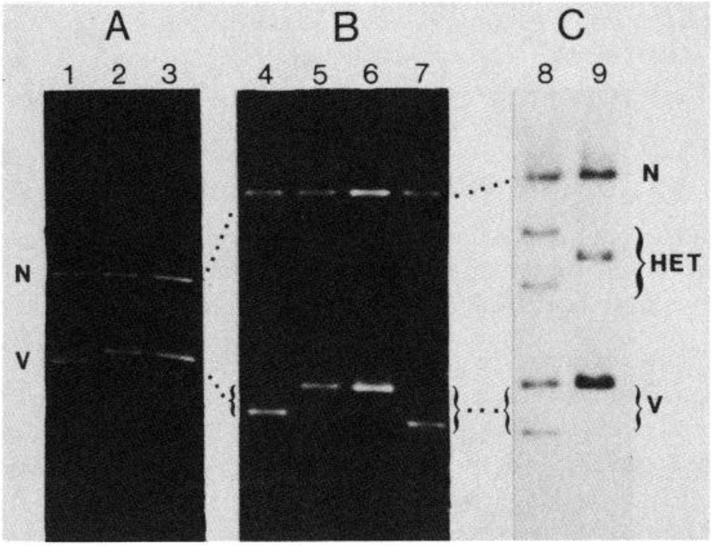


Fig. 2. Parallel DGGE analysis of cloned mouse *adenine phosphoribosyltransferase (aprt)* fragments (parts modified from *[7]*). Variant fragments (V) (560 bp) and 428 bp nonvariant fragments (N) were obtained by *Bam*H1 plus *Xho*1 double digestion of cloned *aprt* DNA as previously described *(7)*. A: Portion of an ethidium bromide stained, 40–90% gradient gel. Lanes 1, 2, and 3 illustrate V fragments that differ from one another at the same site. Lane 1: wild-type (contains a G at the variable site); lane 2: mutation to A; lane 3: mutation to T. Note that the V fragments in lanes 2 and 3 are indistinguishable. B: Portion of an ethidium bromide stained 50–70% gradient gel demonstrating the increased band separation that can be achieved by narrowing the gradient range. All possible base substitutions at one site are shown for the V fragment. Lanes 4, 5, and 6 contain the same samples as 1, 2, and 3, respectively. Lane 7 has a G to C point mutation at the site. Note that the V fragments in lanes 5 and 6 are still indistinguishable, even though the gradient range has been changed. C: Probed electroblot demonstrating the heteroduplexes (HET) that were produced following denaturation and annealing of a mixture of DNAs that differ in sequence. Lane 8: Mixture of DNA samples from lanes 6 and 7, denatured, and allowed to anneal prior to DGGE. Lane 9: Mixture of DNA samples from lanes 5 and 6, denatured, and allowed to anneal prior to DGGE. In any mixture of two different DNA samples, two different heteroduplexes will be produced. The two heteroduplexes produced in lane 8 are evident. Two heteroduplexes also are produced in lane 9, but they both denature at the same position in the gradient, producing only a single visible band *(7)*.

Once the approximate denaturing concentration has been established for a particular fragment(s), the extremes of the gradient should be narrowed. This will increase separation between fragments (cf Figs. 2A and 2B). For best results, use a gradient in which the fragment(s) of interest will denature approximately halfway down the gel. For example, if broad-

range parallel or perpendicular gel analysis indicates that a particular fragment denatures at approx 60% denaturant, for subsequent analysis of the fragment, change the gradient to 50–70%.

3. Heteroduplex formation: Some DNA fragments that differ from each other by a single base pair cannot be distinguished by DGGE, even if the sequence difference resides within an early melting domain (cf lanes 5 and 6 in Fig. 2B). Such single base-pair differences, however, can frequently be resolved following formation of heteroduplexes.

   When a mixture of two DNA fragments that differ at one or more base pairs is denatured and allowed to anneal, the products will include each of the parental homoduplexes and two heteroduplexes that contain a region(s) of mismatch. Such mismatches destabilize the helix and result in a reduced melting temperature compared to that of each of the homoduplexes. As a consequence, the heteroduplexes melt at a lower concentration of denaturant than the homoduplexes (Fig. 2C). Detection of heteroduplexes during DGGE establishes that one or more sequence differences probably exist between the parental homoduplexes. Thus, analysis of heteroduplexes can reveal sequence differences that are cryptic in the analysis of homoduplexes (e.g., *see* Fig. 2, lanes 5, 6, and 9). As a model, we have analyzed all permutations of base substitutions at the terminal 2 nucleotides of the splice acceptor site in the third intron of the mouse *adenine phosphoribosyltransferase (aprt)* gene.

   All possible transitions and transversions in DNA at that site can be detected by DGGE analysis of heteroduplexes *(7)*.

   Heteroduplexes are produced by mixing the DNA fragments of interest in high salt buffer (300 m*M* NaCl, 1 m*M* EDTA, 30 m*M* Tris-HCl, pH 8), heating to 98°C for 10 min, and allowing annealing to occur at 65°C (for 3–4 h). After annealing, 2.5 vol of ethanol are added directly to the mixture (no additional salt is required), the DNA is precipitated at –20°C or –70°C, and resuspended in appropriate buffer for further restriction enzyme digestion or DGGE.

4. GC clamping: This approach can be used to increase the efficiency of mutation detection by DGGE *(5,6,13–15)*. A GC-rich region termed a "GC clamp" (100% intermixed G + C) is added to one end of the DNA fragment. Because the clamp is very thermostable, it serves as the last melting domain. This enables mutation detection in the segment of a DNA fragment that was previously a late melting domain. For example, Myers et al. *(6,14)* have shown that attachment of a GC clamp to a fragment of the beta globin promoter enables DGGE detection of mutations in two melting domains of the fragment whereas, without the clamp,

DGGE can only detect mutation in the first melting domain. For analysis of cloned DNA, GC clamps can be directly ligated to the fragment(s) *(14)*. For analysis of PCR-amplified DNA, a convenient way to produce GC-clamped fragments is by inclusion of a 40–45-bp G + C rich sequence at the 5' end of one or both of the two primers *(16)*. GC clamps on both primers can be used if a PCR fragment is to be cut by restriction enzyme digestion prior to DGGE.

3.2. Sample Preparation

Fragments should be between 50 and 1000 bp, so that only one or a few melting domains are present. This is most easily accomplished by restriction endonuclease digestion of the DNA. If the sequence or a restriction map of the DNA is available, one or a combination of enzymes that will yield fragments within this size range is chosen. If no such information is available, use restriction enzymes that cut often, such as *Hae*III, *Alu*I, *Msp*I, *Taq*I, and so on, in order to generate small fragments. Use of different enzymes, in separate experiments, is advantageous, since sequence context and melting domains will be altered. Thus, one may increase the efficiency of detection of sequence changes *(4,7)*.

The optimum amount of DNA loaded on the gel will vary, depending on the experiment. For cloned DNA, 200 ng of a single fragment is readily visible in ethidium bromide stained denaturing gels. Much less DNA (cloned) can be used if the DNA is transferred (Southern blotted) to a solid matrix, such as Nytran, and hybridized with a radiolabeled probe, or if the fragments themselves are radioactively labeled.

In some studies, denatured genomic DNA fragments have been analyzed following hydridization to cloned, radiolabeled single-strand DNA with subsequent DGGE and autoradiography *(8,9)*. Alternatively, genomic DNA fragments have been blotted onto a membrane after DGGE fractionation, and analyzed by hybridization with radiolabeled probes (Mark Gray, personal communication). A more convenient method, however, entails the analysis of genomic fragments that have been amplified by PCR.

3.3. Amplification of Samples by PCR

Either cloned or genomic DNA amplified by PCR is suitable for DGGE (Fig. 3). This approach is particularly advantageous for analysis of mutations in genomic DNA because PCR amplified fragments can be visualized in ethidium bromide-stained denaturing gels, thus obviating the need for blotting and hybridization with radioactive probes. Standard methods for performing PCR should produce usable fragments. If PCR amplification is used, one should select conditions that result in the production of a single fragment (i.e., use 20–30-mer primers and high annealing and extension tem-

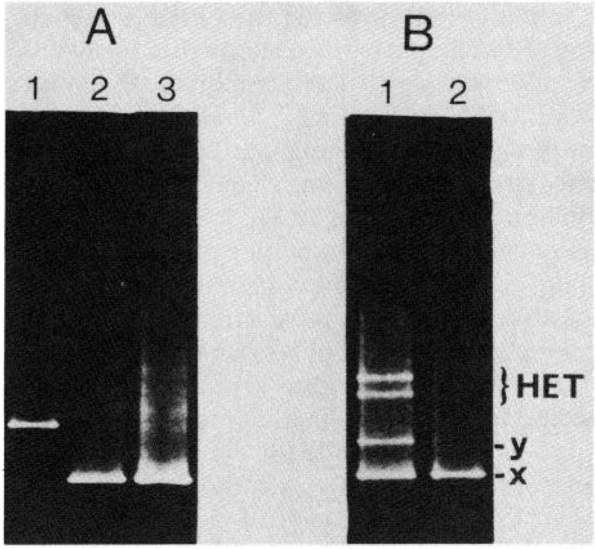


Fig. 3. Parallel DGGE analysis of the mouse *aprt* V fragment obtained from PCR-amplified DNA. In these analyses, the V fragment (560 bp) is homologous to that shown in Fig. 2 because it was obtained by restriction endonuclease digestion (*Bam*HI plus *Xho*I) of a larger fragment (773 bp) that was amplified. A: DGGE analysis of DNA that has been PCR-amplified from dilute plasmid DNA preparations (ca. 5 ng). Lane 1: Wild-type V fragment; lanes 2 and 3: V fragment with point mutation (A to G) at the nucleotide adjacent to that contained in the DNA shown in Fig. 2. B: DGGE analysis of DNA that has been PCR amplified from genomic DNA. Genomic DNA was isolated from a human cell line (HTD114 derivative) that contains two copies of a mouse *aprt* transgene. Lane 1: DNA from a clone derived after ethylmethanesulfonate (EMS) treatment of the cell line. Lane 2: DNA from the cell line prior to EMS treatment. Samples for both lanes were denatured and allowed to anneal prior to DGGE. Lane 1 demonstrates DGGE detection of an EMS-induced mutation (actually a reverse mutation) in the V fragment of one of the two mutant *aprt* transgenes that are present in this cell line. HET: heteroduplexes; X homoduplex of mutant V fragment; Y: homoduplex of wild-type (reverted) V fragment.

peratures). This will help avoid confounding multiple bands in the DGGE gel. Even when a single fragment has been amplified, minor bands or smears sometimes are seen with DGGE analysis of PCR-amplified DNA (*see* Fig. 3A, lane 3). This is owing, in part, to errors introduced during amplification. In fact, DGGE has been used as a method to assess the error rate of different polymerases that may be used for PCR *(15)*.

To analyze a large gene or DNA region by PCR amplification and DGGE, a set of appropriate small fragments must be produced. This can be accomplished with multiple sets of primers to divide the region into a series of smaller

fragments for amplification, or by the use of restriction enzymes to cleave large PCR-amplified fragments. Thus, although it is not always necessary to use restriction endonucleases to digest amplified DNA prior to DGGE, such digestion is useful (i) if long sequences are amplified, (ii) if one is interested in the behavior of a particular subfragment, or (iii) if it is necessary to alter general sequence context and melting domains (*see* Section 3.2).

Figure 3B illustrates the use of DGGE for the detection of a single-base-change in a fragment produced by amplification of genomic DNA. In this example, the target of amplification is a mouse *aprt* transgene, two identical copies of which have been stably transfected into a human cell line. Thus, only a single band is seen on a DGGE gel (lane 2). However, a more complex pattern (lane 1) with multiple bands is shown for a cell clone derived after treatment with the mutagen ethylmethanesulfonate (EMS). In this clone, there are now two different copies of the mouse *aprt* gene (fragments x and y), and heteroduplexes (HET) are seen. Although transgenes are analyzed in this example, lane 2 illustrates the type of pattern that is also seen for individuals who are homozygous for a particular region or "allele" of DNA (no evidence of mutation/sequence differences). Complex, multiple-band patterns (e.g., lane 1) are seen for individuals who are heterozygous for a particular DNA region. As discussed previously, the formation of heteroduplexes is frequently advantageous for the detection of mutation. However, heteroduplexes are sometimes, but not always, spontaneously formed during the amplification procedure. Therefore, to ensure heteroduplex formation, we denature and anneal in high-salt buffer.

3.4. Making a Parallel Gradient Gel

Preparation of a 20–80% gel is described as an example.

1. Thoroughly clean the glass plates, first with soapy water and then with ethanol. Assemble the apparatus according to the manufacturer's protocol. With the Hoefer system, only side spacers (0.75 or 1.5 mm thick) are used. Sealing with grease to avoid leaks is usually not needed except at the bottom corners where there is contact between the plates and the gel-stand gasket. Gel vol is determined empirically. If leakage is a problem, the outside edges of the plates can be sealed with agarose prior to pouring the gel.
2. Bottom plug gel: A narrow bottom plug gel is poured (Fig. 4). The plug can be acrylamide or agarose. To form an acrylamide plug (1.5–2.5 cm high), prepare an acrylamide solution (e.g., 10% v/v acrylamide stock, 90% v/v denaturant stock) and add appropriate volumes of catalysts (*see* Step 9, this section). Apply from the top of the clamped plates along one edge using a syringe and allowed to flow to the bottom and polymerize.

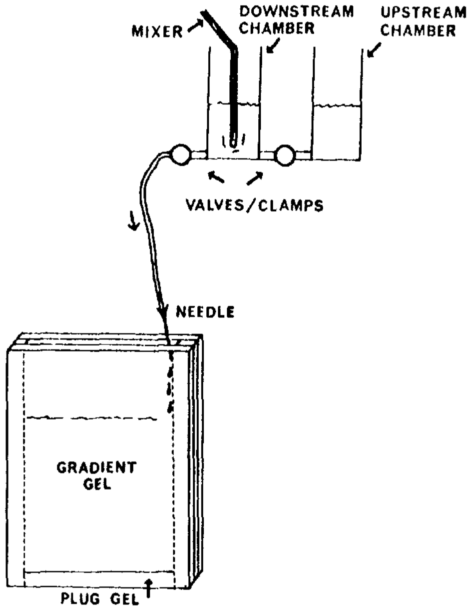


Fig. 4. Schematic of arrangement for pouring gradient gels *(see text for details).* (not drawn to scale)

No overlay solution is needed. Alternatively, melted agarose (1.5% in 1× TAE) can be allowed to rise between the plates from below by capillary action while the plates sit in a trough.

3. Denaturing gradient gel: The setup for making gradients is shown in Fig. 4. Each chamber of the gradient maker should contain approx one-half of the vol needed to fill the gel plate assembly. The upstream chamber contains the solution with the low denaturant concentration and the downstream chamber contains the solution with the high denaturant concentration. The denaturant concentration in each chamber is chosen to produce the desired gradient (*see* discussion in Section 3.1, Step 2).
4. Check polymerization of the plug gel.
5. Prepare an 80% denaturant solution: Mix 9.6 mL of 100% denaturing stock solution and 2.4 mL of nondenaturing stock solution in a small glass beaker on ice.
6. Prepare a 20% denaturant solution: Mix 2.4 mL of 100% denaturing stock solution and 9.6 mL of nondenaturing stock solution in a small glass beaker on ice.
7. Attach tubing from the gradient maker to the gel plate assembly and close both valves/clamps.

8. Set up a magnetic stirrer or vibrating mixer for the downstream side of the gradient maker.
9. Add catalyst (60 µL of 10% APER and 5 µL of TEMED) to the high concentration (80%) denaturant solution in the beaker on ice, mix well by swirling, and pour into the downstream side of the gradient maker.
10. To remove air bubbles, open the valve that separates the two chambers of the gradient maker and allow a small amount of the high concentration denaturant solution to flow into the upstream chamber; then, close the valve. If excess high concentration denaturant gets into the upstream chamber, return it to the downstream chamber with a pipet.
11. Be certain that the mixing device is operating in the downstream chamber.
12. Add catalyst (as in Step 9) to the low concentration denaturant solution, mix well, and pour into the upstream chamber.
13. Open the valve between the chambers.
14. Open the valve or clamp between the downstream chamber and the gel plate assembly. The solution can be delivered to the gel by gravity or peristaltic pump. Fill slowly to the top of the plates. Filling times between 10 and 30 min work well.
15. Insert a slot former (10–20 slot comb) immediately after pouring the gel. Polymerization takes between 15 and 45 min. An alternative to placing the comb directly into the gradient gel is to pour the gradient (as in Step 14), but leave about a 2–2.5-cm space above and overlay the gel solution with a small amount of butanol. After the gradient has polymerized, an upper gel (0% denaturant) can be poured and the comb inserted.
16. The gel can be run immediately after polymerization or it can be left covered with plastic wrap overnight at room temperature.

3.5. Electrophoresis

1. Temperature equilibration of chamber: At least 1 h prior to use, electrophoresis buffer is put into the lower buffer chamber, a stir bar placed in the chamber, the magnetic stirrer turned on, and a thermometer placed in the chamber. The circulating heating bath is turned on and set at a temperature sufficient to heat the electrophoresis chamber to 60°C (determined empirically). For example, a circulating bath temperature of 63.5°C is required for a Hoefer SE660 (insulated) Neslab RT3-210D combination.
2. Loading the gel and starting electrophoresis: After the comb has been removed, the wells should be rinsed (1–2×) with electrophoresis buffer,

and the samples mixed with loading buffer (1× final concentration). The gel can be loaded while in the gel-pouring stand by using an adjustable micropipet or a microsyringe. If a microsyringe is used, it should be rinsed three times between samples. Once loaded, the gel and upper buffer chamber assembly are placed into the lower chamber and buffer is added to the upper chamber. If a microsyringe is used to load samples, the gel can be loaded after it has been put into the electrophoresis chamber. This is useful for temporally staggered loading of samples once electrophoresis has begun. Once the loaded gel is in place, a cover is placed over the apparatus to reduce evaporation and heat loss and the buffer recirculation pump started. If the gel is loaded while outside the chamber, it will be cool and the temperature of the chamber will drop slightly when the gel is placed inside. Chamber temperature usually recovers quickly to 60°C. For many analyses, electrophoreses can be started before the chamber temperature equilibrates to 60°C.

Checklist:

a. _____Gel loaded and in place.
b. _____Buffer recirculating pump on.
c. _____Heat source (circulating water bath) on.
d. _____Magnetic stirrer in chamber working.
e. _____Thermometer in chamber

3. Gel electrophoresis: Parameters for electrophoresis depend on the fragment(s) being analyzed and the gradient used (*see* Section 3.1.2). Voltages can range between 60 and 160, and run times between 6 and 24 h (*see* Note 3). These should be determined empirically (*see* Note 5). Because fragment mobility is significantly retarded following domain denaturation, moderately extended run times generally are acceptable because the denatured fragments remain in relatively fixed positions in the gel and separations between fragments are maintained.
4. Postelectrophoresis: The gel is stained with 0.5 µg/mL ethidium bromide *(17)* and the DNA bands visualized on a UV transilluminator (*see* Notes 2 and 4). DGGE gels can also be electroblotted onto nylon membranes, such as Nytran, and hybridized with radiolabeled probe.

4. Notes

1. Perpendicular gel electrophoresis: In this configuration, the denaturant concentration gradient is perpendicular to the direction of electrophoresis *(2,6,14)*. This enables one to more easily determine the position in the gradient at which denaturation occurs. A single sample, containing

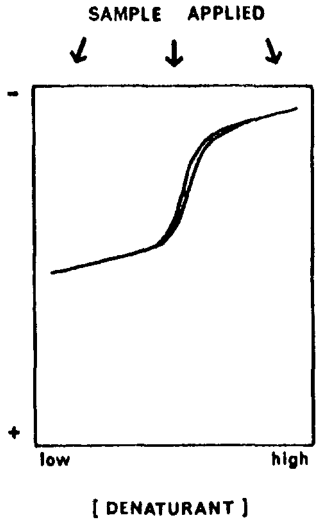


Fig. 5. Diagrammatic representation of the results of a perpendicular gradient analysis of a mixture of two DNA fragments. The splitting of the band at the inflection of the curve indicates a sequence difference between the two samples. The position of the inflection indicates the denaturant concentration where melting occurs.

one or a mixture of DNA fragments, is applied across the top of the gel. After electrophoresis, the DNA forms an S-shaped curve in which the inflection of the curve indicates the point at which denaturation is occurring (Fig. 5). Two fragments that are identical in size but differ in sequence in an early melting domain will comigrate in low concentration denaturant because the DNA migrates only according to molecular size. However, as the fragments encounter higher denaturant concentrations, they separate (note the split in the curve in Fig. 5). Because fragments migrate according to molecular size on the low denaturant side of the gel, a mixture of different sized fragments can be analyzed simultaneously in a single perpendicular gel and the denaturation points determined for each fragment

In preparing a perpendicular gradient, the glass plates should be assembled so that the gel can be poured with the plates on their side. With the Hoefer SE 660, this can be accomplished by using the same plate assembly apparatus as for a parallel gel, except that an additional spacer, slightly shorter than the width of the gel, is inserted between the plates. Assemble the glass plates and spacers as for a parallel gel, but place the

additional spacer (lightly greased) between the top of the plates and abutting the right spacer such that the upper right corner will be leakproof. This will leave a small gap in the upper left corner of the frame. Clamp the top edge with a binder clip. Turn the plates clockwise 90°. Pour the gradient through the gap as described previously. No comb is inserted. Once the gel is polymerized, the gel is rotated back to the vertical position and will have a gap along one edge. Prior to loading the sample, the gap should be filled with an agarose or acrylamide gel, or otherwise sealed. The top spacer is removed, the DNA sample (100–200 µL) is loaded along the entire top edge (*see* Fig. 5), the apparatus is assembled, and electrophoresis is initiated.

2. Samples in lanes at the sides of DGGE gels usually have curved bands after denaturation.
3. The bromophenol blue marker dye usually runs off of the gel before the run is finished.
4. Some DNA fragments do not appear as discrete bands, but rather as streaks or smears because of complex denaturation behavior.
5. Inclusion of ethidium bromide in the gel is not useful for evaluating the status of a run.
6. Methylation can alter DNA melting behavior *(6)*. This is important if one tries to compare PCR-amplified DNA with cloned DNA that has been derived from a source (e.g., bacterial strain) capable of methylation (our unpublished results).

Acknowledgments

This work was supported by National Institutes of Health grants DK38185 and CA36897 to J. A. Tischfield and P. J. Stambrook.

References

1 Fischer, S. G. and Lerman, L. S. (1979) Length-independent separation of DNA restriction fragments in two-dimensional gel electrophoresis. *Cell* **16,** 191–200.

2. Myers, R. M , Maniatis, T , and Lerman, L. S. (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods Enzymol* **155,** 501–527

3 Lerman, L S. and Silverstein, K. (1987) Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Methods Enzymol* **155,** 482–501

4 Lerman, L. S , Silverstein, K., and Grinfeld, E. (1986) Searching for gene defects by denaturing gradient gel electrophoresis. *Cold Spring Harbor Symp. Quant Biol* **51,** 285–297

5. Fischer, S G. and Lerman, L. S. (1983) DNA fragments differing by single-base pair substitutions are separated in denaturing gradient gels: Correspondence with melting theory. *Proc. Natl. Acad Sci. USA* **80,** 1579–1583.
6. Myers, R. M., Fischer, S G., Lerman, L S., and Maniatis, T. (1985) Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* **13,** 3131–3146
7. Dlouhy, S. R., Schaff, D. A., Trofatter, J. A., Liu, H. S., Stambrook, P. J., and Tischfield, J A. (1989) Denaturing gradient gel analysis of single-base substitutions at a mouse adenine phosphoribosyltransferase splice acceptor site. *Mol. Carcinogenesis* **2,** 217–225.
8. Myers, R. M., Lumelsky, N., Lerman, L S., and Maniatis, T. (1985) Detection of single base substitutions in total genomic DNA *Nature* **313,** 495–498
9 Noll, W. W and Collins, M (1987) Detection of human DNA polymorphisms with a simplified denaturing gradient gel electrophoresis technique. *Proc. Natl. Acad. Sci USA* **84,** 3339–3343
10 Cariello, N. F., Scott, J K., Kat, A G., Thilly, W. G., and Keohavong, P (1988) Resolution of a missense mutant in human genomic DNA by denaturing gradient gel electrophoresis and direct sequencing using in vitro DNA amplification: $HPRT_{Munich}$ *Am. J. Hum. Genet.* **42,** 726–734.
11. Amselem, S., Duquesnoy, B. S., Attree, O , Novelli, G , Bousnina, S., Postel-Vinay, M-C , and Goossens, M. (1989) Laron dwarfism and mutations of the growth hormone-receptor gene. *N. Engl J. Med.* **321,** 989–995.
12. Theophilus, B. D M., Latha, T., Grabowsky, G. A , and Smith, F. I. (1989) Comparison of RNase A, a chemical cleavage and GC-clamped denaturing gradient gel electrophoresis for the detection of mutations in exon 9 of the human acid β-glucosidase gene. *Nucleic Acids Res.* **17,** 7707–7722.
13. Sheffield, V. C., Cox, D. R., Lerman, L. S., and Myers, R. M (1989) Attachment of a 40-base-pair G + C rich sequence (GC clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes *Proc. Natl. Acad Sci. USA* **86,** 232–236.
14 Myers, R. M , Fischer, S. G., Maniatis, T., and Lerman, L. S. (1985) Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis *Nucleic Acids Res.* **13,** 3111–3129.
15. Keohavong, P. and Thilly, W. (1989) Fidelity of DNA polymerases in DNA amplification. *Proc Natl. Acad. Sci USA* **86,** 9253–9275.
16 Myers, R. M., Sheffield, V. C., and Cox, D R. (1989) Mutation detection by PCR, GC-clamps, and denaturing gradient gel electrophoresis, in *PCR Technology* (Erlich, H A., ed.) Stockton, NY, pp 71–88
17. Maniatis, T., Fritsch, E. F , and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 173–177

Chapter 11

The Detection and Mapping of Point Mutations by RNase A Cleavage

J. Ross Hawkins and Raymond Dalgleish

1. Introduction

Until recently, most defects identified in mutant genes have been based on large size differences, as detected by Southern or Northern blotting, or by the sequencing of cloned DNA fragments. However, it is probable that a large proportion of disease-causing mutations are point mutations. As a consequence of the rarity of RFLP markers showing linkage disequilibrium with mutant genes, and of point mutations that create or destroy restriction endonuclease recognition sites, it is of fundamental importance in medical molecular genetics to have a technique that will allow the direct detection of point mutations. Several techniques have been developed that go some way toward this goal: most notably, denaturing gradient gel electrophoresis *(1–3),* Ribonuclease A (RNase A) cleavage *(4–6),* and more recently, chemical cleavage *(7,8).*

The technique of RNase A cleavage allows the detection and mapping of point mutations and polymorphisms in DNA or RNA. In this chapter, the technique of RNase A cleavage will be described with specific reference to the mapping of mutations in RNA transcripts (*see* Figs. 1 and 2). A summary of denaturing gradient gel electrophoresis and chemical cleavage will be presented later, along with the potential contribution of the polymerase chain reaction to these techniques.

The use of RNA as the substrate has the advantage that only the coding-region of the gene of interest is examined, and the presence of the poly(A)

From: *Methods in Molecular Biology, Vol 9. Protocols in Human Molecular Genetics*
Edited by. C. Mathew Copyright © 1991 The Humana Press Inc , Clifton, NJ

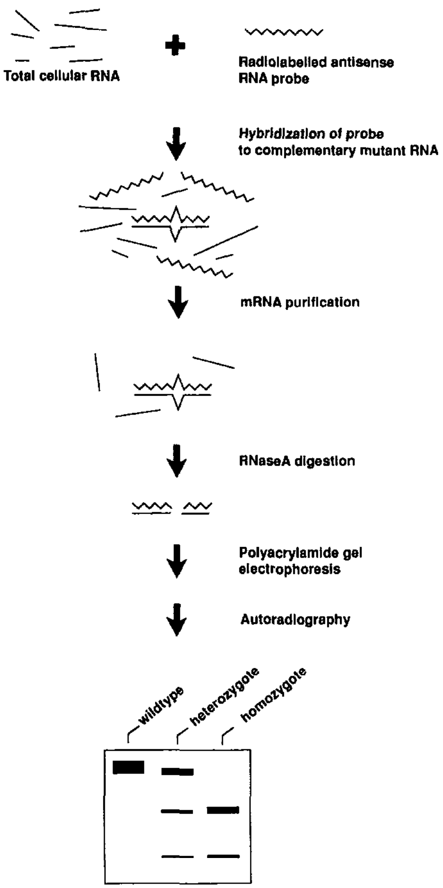


Fig. 1. Diagrammatic representation of the RNase A cleavage methodology. A single-stranded "antisense" RNA probe of uniform length is transcribed in vitro from a recombinant plasmid DNA template. The probe is annealed to its complementary mRNA in a liquid hybridization reaction. Unbound probe molecules, along with ribosomal and transfer RNAs are removed in an mRNA purification step. If a mutation is present in the mRNA of interest, a mismatch will exist within the probe/mRNA duplex. Treatment with RNase A digests all single-stranded mRNA, vector sequences away from the ends of the bound probe, and any mismatched sequences in the probe/mRNA duplex. The duplex RNA is denatured and run on a denaturing (sequencing-type) polyacrylamide gel. The autoradiogram of the dried gel reveals the size of the "protected" probe molecules. If a mutation is present in the mRNA of interest, two cleavage products of the probe will be present on the autoradiogram. If cleavage of the mismatch is complete, in heterozygous individuals, the fully protected probe and the two cleavage products will be present, whereas in homozygous individuals only the two cleavage products will be present.

Fig. 2. Autoradiogram showing a heterozygous point mutation in α1(I) collagen mRNA. Track A contains protected normal placental RNA and track B contains protected mutant fibroblast RNA. The mutation is a T for G substitution *(10)*. The protected probe is 534 bases in length and the cleavage products are 304 and 230 bases in length.

tail at the 3' end of the mRNA enables a purification step to be included in the procedure.

The technique is capable of detecting mutations in transcripts as low in abundance as 0.01% of the mRNA *(6)*, and can map the position of mutations with an accuracy of about 10 bp. However, the utility of the technique is limited because it is only capable of detecting 30–50% of point mutations.

The only prerequisite for laboratories wishing to use the technique is a wild-type cDNA clone of the gene of interest. The steps involved in this method can be summarized as follows:

1. Subcloning of cDNA fragments into transcription vectors.
2. Synthesis of an antisense RNA probe.
3. Hybridization of probe to complementary mRNA.
4. Purification of mRNA.
5. Digestion with RNase A.
6. Analysis of 'protected' mRNA by polyacrylamide gel electrophoresis followed by autoradiography.

2. Materials

1. Polyacrylamide gel equipment e.g., Bio-Rad Protean II or Bio-Rad Sequi-Gen.
2. In vitro transcription vectors e.g., pTZ18R or pTZ19R. (Pharmacia).
3. Nuclease-free bovine serum albumin (BSA) (Sigma).
4. RNase inhibitor (RNasin) (Promega).
5. T7 RNA polymerase (Boehringer).
6. ^{32}P-CTP (Amersham) (*see* Note 3).
7. RNase-free DNase I (Pharmacia).
8. Messenger affinity paper (mAP) (Amersham).
9. Phenol/chloroform/isoamyl alcohol/8-hydroxyquinoline (100:100:4:0.1; v:v:v:w; equilibrated with 10 m*M* Tris-HCl, pH 7.6).
10. Chloroform/isoamyl alcohol (24:1; v:v).
11. Diethyl pyrocarbonate (DEPC).
12. 2*M* Sodium acetate, pH 5.6.
13. TE: 10 m*M* Tris-HCl, 1 m*M* EDTA, pH 7.6.
14. 10× NTP mix: 5 m*M* GTP, 5 m*M* ATP, 5 m*M* UTP, 75 µ*M* CTP.
15. 5× probe synthesis buffer: 200 m*M* Tris-HCl, pH 8.25; 30 m*M* $MgCl_2$; 10 m*M* spermidine.
16. 100 m*M* Dithiothreitol (DTT).
17. 7.5*M* Ammonium acetate.
18. Deionized formamide.
19. Hybridization buffer: 2*M* NaCl, 0.2*M* PIPES, pH 6.7.
20. 0.5*M* NaCl.
21. 2.5*M* NaCl.
22. Ribonuclease A (RNase A): Dissolve at 5–100 µg/mL in 200 m*M* NaCl, 100 m*M* LiCl, 30 m*M* Tris-HCl, 3 m*M* EDTA, pH 7.5, and boil for 10 min.
23. 10% Sodium dodecyl sulfate (SDS).
24. 10 mg/mL Proteinase K.
25. 10 mg/mL Transfer RNA (tRNA).
26. 10× Tris-Borate-EDTA (TBE): 0.89*M* Tris base, 0.89*M* boric acid, 30 m*M* EDTA.

27. Polyacrylamide gel: 4.75% Acrylamide, 0.25% *bis*-acrylamide, 7*M* urea, 20% formamide, 0.5 × TBE.
28. 10% Ammonium persulfate.
29. Tetramethylethylene diamine (TEMED).
30. Electrophoresis tracking dye: 1 mg/mL Xylene cyanol FF, 1 mg/mL bromophenol blue, 10 m*M* EDTA in deionized formamide.

3. Methods

3.1. Preparation of Probe Templates

1. Subclone cDNA fragments into either of the transcription vectors pTZ18R or pTZ19R, in such an orientation as to enable the in vitro synthesis of 'antisense RNA' (*see* Note 15).
2. Purify recombinant plasmid DNA by cesium chloride gradient centrifugation.
3. Linearize 50 µg DNA at the 5' end of the insert (with respect to its in vivo arrangement) with a restriction endonuclease generating a 5' overhang or blunt end *(9)*.
4. Extract the DNA with phenol/chloroform three times and remove traces of phenol by extracting with chloroform/isoamyl alcohol. All handling from this point onward is as for RNA (i.e., with DEPC-treated solutions and glassware).
5. Precipitate the DNA with a 0.25 vol of 2*M* sodium acetate, pH 5.6 and 2.5 vol of ethanol. Rinse the pellet with 70% ethanol, drain, and vacuum dry.
6. Resuspend the pellet initially in 20 µL of TE, quantify, and adjust to 0.4 µg/µL.

3.2. Antisense RNA Probe Synthesis (see Note 2)

1. Prepare the following reaction mix:

10× NTP mix	1 µL
100 m*M* DTT	1 µL
5× buffer	2 µL
2 mg/mL BSA	0.5 µL
RNasin	— (35 U)
T7 RNA polymerase	— (10 U)
^{32}P-CTP (800 Ci/mmol)	— (40 µCi)
Template DNA	0.6 µL (0.25 µg)
Distilled water	—
	10 µL

The reaction mix is incubated at 37°C for 2 h.

2. Add 6 U of DNase I and return the reaction mix to 37°C for a further 15 min (*see* Note 5).
3. Return the tube to room temperature, add 40 μL of distilled water and phenol extract.
4. Transfer the aqueous (upper) layer to a fresh tube and reextract the organic layer with 50 μL distilled water. Pool the aqueous phases.
5. Extract with chloroform/isoamyl alcohol to ensure removal of any contaminating phenol.
6. Precipitate the aqueous phase with 0.5 vol of 7.5*M* ammonium acetate and 3.4 vol of ethanol. Incubate the tube at –80°C for 15 min and centrifuge for 15 min.
7. Remove the supernatant with a micropipet, respin the tube briefly, and remove any remaining ethanol again with a micropipet. Do not vacuum-dry the pellet as resuspension will be very difficult.
8. Resuspend the (invisible) pellet in 80 μL of distilled water. Resuspension will require pipetting the sample up and down many times. Determine the extent of resuspension by monitoring with a Geiger counter. Make sure the pellet is at least 40% resuspended (*see* Note 7).

3.3. Hybridization

1. To 20–200 μg of total cellular RNA, add 10^5–10^6-cpm probe, adjust to 0.1*M* NaCl, and add 2.5 vol of ethanol.
2. Mix and leave on ice for 5 min before centrifugation.
3. Drain the tube as before.
4. Resuspend the pellet in 160 μL of good-quality deionized formamide. This is best done by vigorously squirting the formamide onto the pellet, such that the pellet will be dislodged from the side of the tube. Then the pellet should float and dissolve very quickly. Leave the sample at room temperature for 30 min.
5. Make sure that the pellet is fully dissolved, and add 40 μL of 5× hybridization buffer.
6. Heat the sample at 85°C for 5 min to remove secondary structures from the RNA.
7. Incubate at 60°C overnight for hybridization.

3.4. mRNA Purification

1. With gloved hands and a sterile scalpel blade, cut a piece of Amersham mAP (messenger affinity paper) approx 4 mm × 4 mm. Try to avoid

touching the paper (*see* Note 11). Transfer it, using a sterile syringe needle, to a small Petri dish containing 0.5*M* NaCl. Allow to soak for 5 min.

2. Retrieve the hybridization tube and add to it 100 µL 0.5*M* NaCl. Mix by pipetting.
3. Place a piece of Whatman 3 MM paper (4 cm × 4 cm, autoclaved, and dried) behind a radioactivity shielding screen. Place the mAP onto the 3 MM paper. With a micropipet, set to 10 µL, spot the RNA onto the mAP drop by drop. This should be done slowly and carefully, making sure the RNA soaks through the mAP and is not directly absorbed by the 3 MM paper. Initially, the mAP will take up the RNA rapidly, however, with each drop applied, the time taken will increase.
4. When all the RNA has been loaded, transfer the mAP to a Petri dish containing 0.5*M* NaCl. Leave for approx 15 min, agitating occasionally. Repeat this step twice with fresh solutions of 0.5*M* NaCl.
5. Remove the mAP and blot dry by touching the edge of a tissue and place in a fresh microfuge tube.
6. Add 180 µL of distilled water and incubate at 65–70°C for 10 min.
7. Chill the tube on ice for 5–10 min. With a needle, bring the mAP to the top of the tube and trap it in the lid. Centrifuge the tube for a few seconds, then remove and discard the mAP. Greater than 80% of the counts should have been eluted into the water.
8. Transfer the tube to room temperature and add 20 µL of 2.5*M* NaCl.

3.5. RNase A Cleavage

1. Add 100 µL of 5–100 µg/mL RNase A to the mAP eluate, mix by pipetting, and leave at room temperature for 30–45 min (*see* Note 12).
2. Add 200 µL of 10% SDS and 10 µL of 10 mg/mL proteinase K and incubate at 37°C for 20 min.
3. Add 1 µL of 10 mg/mL tRNA and phenol extract.
4. Collect the aqueous phase, taking care not to carry over any phenol.
5. Add 750 µL of ethanol, chill on ice for 5 min, centrifuge for 15 min, and drain.
6. Wash the (invisible) pellet in 200-µL ethanol, centrifuge for 5 min, drain, and *briefly* vacuum-dry. This is the "protected" RNA.

3.6. Polyacrylamide Gel Electrophoresis

1. Clean the gel plates thoroughly and siliconize one plate.
2. Assemble the gel mold and pour the gel. Gels should be between 0.25 and 1mm in thickness.

3. Prerun the gel to allow it to warm. If a gel apparatus is used that contains a water-circulating system (such as, Bio-Rad Protean II), warm water should be circulated to keep the gel at a constant warm temperature (approx 45°C) (*see* Note 13).
4. Resuspend the "protected" RNA in sufficient tracking dye, such that 10–40 cps is loaded onto the gel.
5. Incubate the samples at 90°C for 5–7 min to denature the double-stranded RNA.
6. Following denaturation, load the samples rapidly onto the gel.
7. Run the gel at approx 25 mA for 2–3 h.
8. Disassemble the gel, fix, and dry as for sequencing gels.
9. Autoradiographic exposure overnight should be sufficient. Longer exposures may be required for low-abundance RNAs and to visualize the products of partial RNase cleavage.

4. Notes

1. The T7 RNA polymerase method of in vitro transcription is very efficient and can yield microgram quantities of RNA.
2. It is essential that the probes produced by this technique are full-length. This can be checked by running an aliquot of the probe on the gel.
3. The use of CTP as the labeled nucleotide is not necessary, and many groups, in fact, prefer GTP *(9)*.
4. The probe DNA template must be linearized to restrict transcription to the plasmid insert only. Cotranscription of vector DNA will reduce the relative activity of the probe and increase background.
5. Complete digestion of the template DNA with DNase I following transcription is essential, as any contaminating DNA will interfere with the hybridization of the probe to its target RNA. Template DNA digestion can be checked by performing a nonradioactive in vitro transcription reaction, and checking that the vector DNA band is not visible on an agarose gel.
6. The use of fresh isotope for the probe synthesis is preferable, particularly for low-abundance RNAs.
7. Following the ethanol precipitation of the radioactive probe, resuspension of the counts always requires much pipetting. A significant proportion of the counts always remains stuck to the tube.
8. It is advisable in initial experiments, at least, to measure the activity of the probe by Cerenkov counting an aliquot.
9. Formamide for the hybridization may be sufficiently deionized by successive freezings at 0°C and pouring off any unfrozen liquid. Although

less effective than using deionizing beads, this method prevents the introduction of RNases into the formamide.

10. The purification of mRNA enriches for the target mRNA and also removes unbound probe, allowing for very clean autoradiographs with low background.
11. It is important not to touch the mAP, and manipulations should always be performed by picking the paper up on the end of a sterile needle.
12. The concentration of RNase A used in the cleavage reaction will have to be titrated according to the abundance of the mRNA of interest. The RNase A will tend to degrade the double-stranded RNA, albeit at a low level. Thus, it is of great help when calibrating the system to have one or more known point mutations in the gene of interest to act as "positive controls." Sufficient RNase A should be used to cleave mismatches, but not too much, in order to avoid producing a high background.
13. RNA/RNA duplexes are very stable, thus, it is important that the gel is kept warm to maintain complete denaturation.
14. End-labeled restriction endonuclease digested plasmids act as good size markers. Even when labeled with ^{32}P, a stock kept at –20°C should be good for many weeks.
15. Probes for RNase A cleavage experiments should be in the size range of 350–900 bases. Probes smaller than 350 bases in size are in danger of dissociating from the target mRNA during the elution of the mRNA from the mAP. Probes greater than 900 bases should be avoided if possible for two reasons. First, 900 bases represents about the reasonable upper size limit for resolution on polyacrylamide gels. Second, owing to radiolysis of the high specific activity probes, the background smear of degradation products on the gel reaches an unsatisfactory level with large probes. This smearing is also caused by nonspecific degradation by the RNase A. Hence, the larger the probe, the greater the background. As another consequence of radiolysis, probes should never be stored before use, but used immediately.
16. The extent of cleavage at point mutations is not only dependent on the type of mismatch, but the surrounding sequence as well. The efficiency of cleavage of various different mismatches are discussed in refs. *5* and *11*.
17. In order to map any mutations detected and to confirm the existence of a mutation that gives faint cleavage products, new probes should be synthesized that overlap with or span the previous probe. This way, mismatches can be very finely mapped.

18. Once mapped, the mutation can be rapidly characterized by amplification of the appropriate region by the polymerase chain reaction (PCR), followed by DNA sequencing.

5. Alternative Techniques

5.1. Denaturing Gradient Gels

This is a gel system that exploits the small difference in melting temperature (T_m) of two DNA molecules differing by a single base. Problems are encountered with the detection of single base substitutions in domains of high Tm. This problem has been overcome by the inclusion of a GC-rich region (GC clamp) attached to the end of cloned molecules or PCR-amplified molecules (refs. *12,13,* and Chapter 10).

5.2. Chemical Cleavage

The existence of chemicals that react with mismatched bases in DNA has led to the development of a new technique in mutation detection *(7,8).* The chemicals hydroxylamine and osmium tetroxide react with mismatched C and T bases, respectively, to produce a lesion that is cleavable by an alkali, such as piperidine. The use of sense and antisense probes potentially renders all types of mismatch susceptible to cleavage (*see* Chapters 5 and 6).

Other techniques have been devised for the detection of point mutations *(14,15)* that exploit altered electrophoretic mobilities, but have yet to receive as much attention as the techniques of RNase A cleavage, denaturing gradient gels, and chemical cleavage.

The polymerase chain reaction has a role to play in all these techniques, both in amplifying the target sequence—to increase the sensitivity of the technique used, and to aid characterization of any sequence containing a mutation or base substitution.

Acknowledgments

We thank Richard Gibbs for his help and advice with the RNase A cleavage technology.

References

1 Fischer, S. G. and Lerman, L. S. (1983) DNA fragments differing by single base substitutions are separated by denaturing gradient gels: correspondence with melting theory. *Proc. Natl. Acad. Sci. USA* **80,** 1579–1583.

2. Myers, R. M., Lumelsky, N., Lerman, L. S., and Maniatis, T. (1985) Detection of single base substitutions in total genomic DNA *Nature* **313,** 495–498.

3. Myers, R. M. and Maniatis, T. (1986) Recent advances in the development of methods for detecting single base substitutions associated with human diseases. *Cold Spring Harbor Sym. Quant. Biol.* **LI**, 275–284.
4. Winter, E., Yamamoto, F., Almoguera, C., and Perucho, M. (1985) A method to detect and characterise point mutations in transcribed genes: Amplification and overexpression of the mutant c-Ki-ras allele in human tumor cells *Proc. Natl. Acad. Sci. USA* **82**, 7575–7579.
5. Myers, R. M., Larin, Z., and Maniatis, T. (1985) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**, 1242–1246.
6. Gibbs, R. A. and Caskey, C. T. (1987) Identification and localization of mutations at the Lesch-Nyhan locus by ribonuclease A cleavage. *Science* **236**, 303–305.
7. Cotton, R. G. H., Rodrigues, N. R., and Campbell, R. D. (1988) Reactivity of cytosine and thymine in single base-pair mismatches with hydroxylamine and osmium tetroxide and its applications to the study of mutations. *Proc. Natl. Acad. Sci. USA* **85**, 4397–4401.
8. Grompe, M., Muzny, D. M., and Caskey, C. T. (1989) Scanning detection of mutations in human ornithine transcarbamoylase by chemical mismatch cleavage. *Proc. Natl Acad. Sci. USA* **86**, 5888–5892.
9. Melton, D. A., Krieg, P A., Rebagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984) Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**, 7035–7056.
10. Cohn, D. H., Byers, P. H., Steinmann, B., and Gelinas, R. E. (1986) Lethal osteogenesis imperfecta resulting from a single nucleotide change in one human proα1(I) collagen gene. *Proc. Natl. Acad Sci. USA* **83**, 6045–6047.
11. Lopez-Galindez, C., Lopez, J. A., Melero, J. A., de la Fuente, L., Martinez, C., Ortin, J., and Perucho, M. (1988) Analysis of genetic variability and mapping of point mutations in influenza virus by the RNase A mismatch cleavage method. *Proc. Natl Acad. Sci. USA* **85**, 3522–3526.
12. Myers, R. M, Fischer, S. G., Maniatis, T., and Lerman, L. S. (1985) Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis *Nucleic Acids Res.* **13**, 3111–3130.
13. Sheffield, V. C., Cox, D R., Lerman, L. S., and Myers, R. M. (1989) Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad Sci USA* **86**, 232–236
14. Orita, M, Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms *Proc. Natl. Acad. Sci. USA* **86**, 2766–2770.
15. Kornher, J. S. and Livak, K. J. (1989) Mutation detection using nucleotide analogs that alter electrophoretic mobility. *Nucleic Acids Res* **17**, 7779–7784

Chapter 12

Discontinuous Polyacrylamide Gel Electrophoresis of DNA Fragments

Bruce Budowle and Robert C. Allen

1. Introduction

At present, the most polymorphic genetic markers are DNA regions composed of a variable number of tandem repeats (VNTRs) *(1,2)*. The detection of VNTRs is made possible by restriction fragment-length polymorphism (RFLP) analysis via Southern blotting *(3)*. The length of each DNA fragment is a function of the number of tandem repeats contained within it. The RFLP technique requires at least 10–50 ng of essentially undegraded DNA (10,000–20,000 bp) and isotopic detection to obtain a result; the time required for analysis may be in excess of one week *(4)*.

The amplification of short target-DNA sequences by the polymerase chain reaction (PCR) *(5,6)* also promises to be a useful tool for genetic analysis of biological specimens. PCR is particularly desirable for effecting subsequent analysis from DNA samples of limited quality and quantity, eliminating the need for isotopic detection methods, and reducing the time to obtain results. Further, the sensitivity of detection provided by PCR can be combined with the information content derived from fragment-length polymorphisms of VNTR loci (such polymorphisms have been called AMP-FLPs). A number of VNTR loci that are amenable to amplification by PCR have been identified, and include D17S30 *(7)*, D1S80 *(8)*, and the 3'HVR region of the apolipoprotein B gene *(9)*.

A simple polyacrylamide gel electrophoresis technique for AMP-FLP analysis has been developed, and provides high resolution and increased sen-

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

sitivity of detection compared with agarose gel electrophoresis and ethidium bromide staining, as well as a permanent record. The features of the technique are a discontinuous borate–sulfate buffer system *(10)*, surface loading of samples (no wells are necessary for sample application), and nonisotopic staining using silver *(10, 11)*. Electrophoretic results can be obtained in 2–3 h; resolution is 1–10 bp for fragments up to 1000 bp in length (for separations with a 10-cm interelectrode wick distance) and band widths of 200–400 μ are obtained *(10)*. Resolution potential can be manipulated by changing the ionic strength and viscosity of the resolving gel buffer. This chapter describes a technique that can serve as a guideline to achieving efficient and practical fractionation conditions for AMP-FLPs.

2. Materials

1. Tris-sulfate buffer, pH 9.0, is used to provide the leading sulfate ion. A stock solution of 0.07*M* Tris-sulfate (with respect to sulfate) is prepared using Trizma base, 36.3 g and 1*N* H_2SO_4, 62 mL. Make up to a final vol of 400 mL with distilled water.
2. Tris-borate buffer, pH 9.0, is used to provide the trailing and counter ions. A stock solution of 0.14*M* Tris-borate is prepared using Trizma base, 15.72 g and boric acid, 2.19 g. Make up to a final vol of 250 mL with distilled water.
3. DNA size standards are the 123-bp ladder and the 1-kbase ladder purchased from BRL (Gaithersburg, MD). The standards are diluted 1:10 with Tris-sulfate (using eight parts water and one part Tris-sulfate stock solution). The final concentration of the size standards is 100 ng/μL.
4. Rehyratable polyacrylamide gels (5% T, 3% C; crosslinker *N,N'*-methylenebisacrylamide [BIS]); EC Corporation, St. Petersburg, FL) (for definition of %T and %C, *see* Note 3).
5. Acrylamide stock solution: Acrylamide (Bio-Rad, Richmond, CA), 29.1 g; piperazine diacrylamide (Bio-Rad), 0.9 g. Dissolve in 50 mL of distilled water, filter, and make up to a final vol of 100 mL with distilled water.
6. 20% Glycerol.
7. Ammonium persulfate.
8. Tetraethylmethylene diamine (Temed).
9. Agarose plugs are prepared by boiling 8 g of agarose (ME agarose, FMC, Rockland, ME) in 400 mL of Tris-borate stock solution. The agarose solution is poured into a submarine gel-casting tray designed for the H4 submarine gel-electrophoresis tank from BRL (Gaithersburg, MD) (casting-tray dimensions are 20 × 25 cm). After setting, the gel is sliced

into 1-cm sections (each 20 cm long) and submerged in a storage solution containing Tris-borate stock solution and 0.01% bromophenol blue. The bromophenol blue will diffuse into the agarose plugs and subsequently will serve as a tracking dye to mark the moving sulfate–borate boundary.
10. ICE electrophoresis apparatus (EC Corporation, St. Petersburg, FL).
11. Silicone applicator strip (catalog no. 42989, Serva, Heidelberg, FRG).
12. 10% Ethanol.
13. 1% Nitric acid.
14. 0.012*M* Silver nitrate solution.
15. 0.28*M* Sodium carbonate (anhydrous), 0.019% formalin.
16. 10% Glacial acetic acid.

3. Methods

1. DNA is extracted from whole blood (drawn in EDTA Vacutainer™ tubes) as previously described *(4)*.
2. Amplification of the VNTR loci D1S80 and of the 3' hypervariable region of the apolipoprotein B gene by PCR is performed according to Kasai et al. *(8)* and Boerwinkle et al. *(9)*, respectively.
3. Preparation of polyacrylamide gels. These can be prepared in two ways. The first is using rehydratable gels *(12,13)* (*see* Note 4). Rehydratable polyacrylamide gels (5% T, 3% C; crosslinker BIS) (EC Corporation, St. Petersburg, FL) are submerged (gel-side down) in a solution containing 0.035*M* Tris-sulfate (a 1:2 dilution of the Tris-sulfate stock solution) and 7.1% glycerol. Rehydration time is 20–30 min. After rehydration, the excess solution on the gel surface is gently wiped off using a Mylar™ film. The gel is ready for electrophoresis. Alternatively, gels can be cast by the flap technique *(14,15)* (Fig. 1). Briefly, a parafilm gasket (four layers thick—approx 400 µm) is placed onto the hydrophilic side of a Mylar™ sheet (Gelbond NF for polyacrylamide gels, FMC, Rockland, ME). The template is laid horizontally on a table top. A gel solution *(see below)* is pipeted onto one end of the template and a clean top plate is placed on the template in the same manner as a cover slip is placed on a microscope slide. Gels are allowed to polymerize for at least 3 h at ambient temperature prior to electrophoresis. Gels cast in this manner are stable for 2–4 wk at ambient temperature.

   The gel solution for gels cast by the flap technique is made of acrylamide stock solution (1.83 mL), 20% glycerol (3.80 mL), Tris-sulfate stock solution (5.00 mL), and ammonium persulfate (7.5 mg). Deaerate

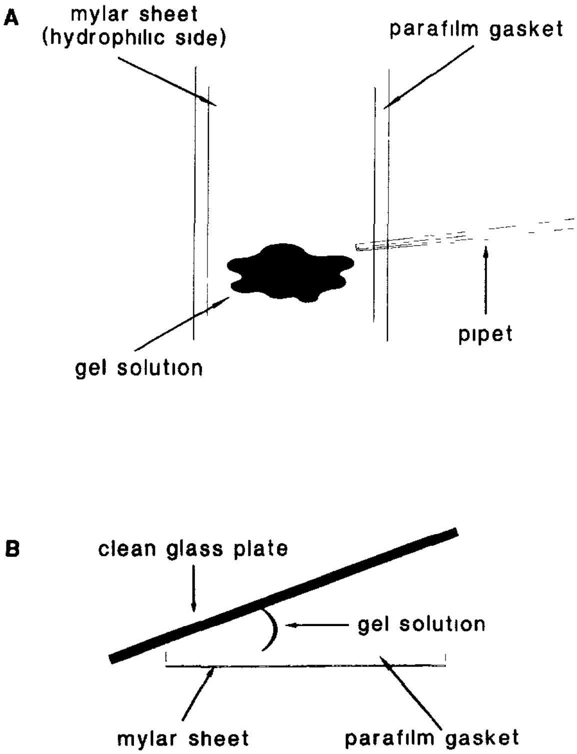


Fig. 1. Display of four-layer-thick parafilm gasket placed on Mylar™ sheet (A). The hydrophilic side of the Mylar™ sheet is face up. The gel solution is poured onto one end of the gel mold (A). A top plate is placed on the gel mold and gel solution (B).

under vacuum for approx 30 s (or until bubbles dissipate). Add TEMED (7.5 µL). Pour the gel immediately.

This gel recipe will prepare a 5% T, 3% C polyacrylamide gel containing 7.1% glycerol and 33 m*M* sulfate. Gels containing different molarities of the stock buffer or varying quantities of glycerol can be made by appropriately adjusting the glycerol and/or Tris-sulfate components.

4. After amplification, the samples are diluted 1:2 with Tris-sulfate (using a 1:5 dilution of Tris-sulfate stock solution).
5. Electrophoretic setup (*see* Fig. 2). Electrophoresis is carried out horizontally at 20°C. The acrylamide gel is placed onto an ICE apparatus (EC Corporation, St. Petersburg, FL). This apparatus is cooled electronically using Peltier units. An agarose plug is placed on each end of the gel. Generally, the distance between the plugs is 10 cm, although this dis-

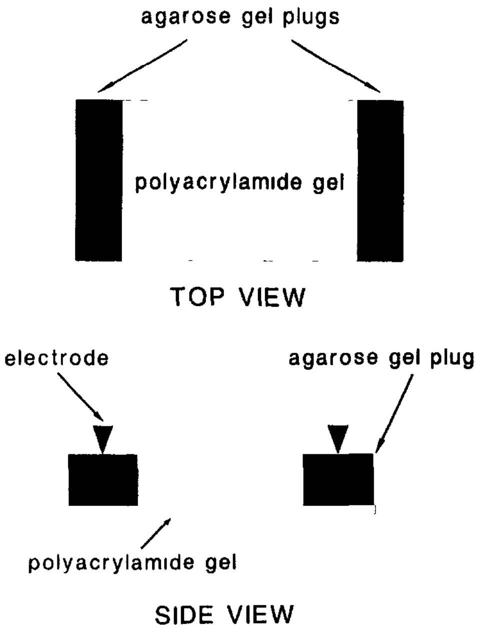


Fig. 2. Display of electrophoretic setup.

tance can be varied according to the demands of the electrophoretic separation. Samples are surface-loaded 1 cm from the cathodal plug. Sample application can be facilitated using a silicone applicator strip (catalog no. 42989, Serva, Heidelberg, FRG). Sample volumes varying from 0.5 to 20 µL have been applied with no apparent adverse effects. The electrodes (the same as those used for isoelectric focusing) are placed on top of the agarose plugs and electrophoretic separation is begun. If the agarose plugs slide, a dry filter-paper strip (Whatman 3MM) can be placed behind the plug to inhibit movement (*see* Fig. 2). Electrophoresis generally is carried out with constant power at 1 W for the first hour, or until the leading edge of the bromophenol blue tracking dye has traveled 1 cm anodal of the sample origin. The power setting is adjusted to 1.5 W for an additional hour, the power setting is then increased to 2.0 W, and electrophoresis is carried out until the leading edge of the bromophenol blue dye has reached the anodal gel-buffer reservoir plug. These power settings are given as a guideline; alternative settings may be employed for the specific needs of the laboratory.

6. Silver Staining.

After electrophoresis, the gels are stained with silver using a modified Merril-stain approach *(11)*. The steps are as follows:

a. Place the gel in a solution containing 10% ethanol for 5 min (gently shake).
b. Oxidize the gel in a 1% nitric acid solution for 3 min.
c. Briefly rinse the gel (a few seconds) in distilled water.
d. Place the gel in a 0.012*M* silver nitrate solution for 20 min.
e. Decant the silver solution, rinse briefly (a few seconds) in water, and reduce the gel in a solution containing 0.28*M* sodium carbonate (anhydrous) and 0.019% formalin. Several changes of the reducing solution may be necessary. The solution should be changed when it turns brown during image reduction.
f. Stop the process with 10% glacial acetic acid (when the desired image has developed).
g. After 5 min, place the gel in distilled water for 5 min.
h. Excess silver deposited on the gel surface can be removed by wiping the surface clean with cotton balls.
i. Air-dry the gel for permanent storage.

4. Notes

1. Polyacrylamide gel is a well-characterized medium on which a variety of electrophoretic processes can be carried out. The gel is the polymerization product of acrylamide monomer and a crosslinking co-monomer (most commonly bisacrylamide). The major advantages of polyacrylamide are that gels can be made in a variety of pore sizes to control sieving effects and that polyacrylamide has little or no electroendosmotic properties and is chemically inert. The material is optically clear and can be stained readily with a variety of available stains. Gels can be covalently bound to both Mylar™ films and glass plates *(10,14,15)*; these supports provide easier handling of polyacrylamide gels, and the support materials are ideal for transmission of densitometric analysis or computer-assisted image capture. Finally, ultrathin-layer gels (<0.5 mm) can be made with polyacrylamide. Compared with thicker gels, ultrathin-layer polyacrylamide gels have a greater surface-area/volume ratio and, thus, are more effective at dissipating Joule heat, require less reagents (i.e., are less costly), and, because of shorter diffusion distances, provide more rapid staining of separated products *(14,16)*. It should be noted that polyacrylamide is a neurotoxin (when unpolymerized), but with proper laboratory care, it is no more dangerous than various commonly used chemicals.

2. Bisacrylamide is routinely used as a crosslinker in polyacrylamide gels. However, piperazine diacrylamide (PDA) can be substituted for BIS on a gram-for-gram basis in the gel stock solution *(17)*. It has been hypothesized that the amide groups of BIS might be partially responsible for background staining when polyacrylamide gels are subjected to silver stain. PDA improves the quality of silver-stained polyacrylamide gels by reducing background staining, the piperazine ring of PDA eliminates hydrogen atoms of amide groups.
3. $\%T = (a + b)/m$; $\%C = b/(a + b)$, where a = grams of acrylamide in gel, b = grams of crosslinker in gel, and m = total volume of gel.
4. Rehydratable polyacrylamide gels *(12,13)* are gels that have been made in Tris-chloride or Tris-sulfate buffer, pH 9.0, and then extensively washed, treated with dextran, and subsequently dried. The washing process removes the buffers, impurities, and unknown polymerization products. The dried gels are essentially empty gels that can be conveniently rehydrated with any buffer. The rehydration is particularly rapid with ultrathin-layer gels, since rehydration equilibration is contingent on gel thickness. Ultrathin-layer gels 350–400 µm thick can be rehydrated in 20–30 min. Dried gels are bound to Mylar™ films or glass to facilitate manipulation and can be stored at ambient temperature for more than a year prior to rehydration without apparent loss of structural or separation functionality.
5. The goal of all electrophoretic methods in sieving media has been to achieve maximal resolution of the components of a complex mixture. The ability to obtain 1–10 bp resolution has obvious advantages for detecting discrete alleles of VNTR loci *(18)*. For such resolution to occur in conventional electrophoresis, the sample components must be concentrated into the narrowest starting zone possible. This will limit the effects of diffusion or resolution of the sample components during electrophoretic separation. Component-zone-sharpening to increase resolution can be obtained in acrylamide gels for both proteins and nucleic acids by a two-step electrochemical procedure described by Allen et al. *(19)*. The sample is treated in such a manner that its ionic strength is one-fifth to one-tenth the ionic strength of the separating gel leading ion concentration. In horizontal gel systems, the sample is placed directly onto the optically flat surface of the gel. The sample components first migrate into the gel at a low voltage by charge and size, and are partially zone-sharpened by the jump from the low sample-buffer conductivity to the higher gel buffer conductivity *(20)*. Next, the high-voltage gradient across the moving boundary sharpens sequentially the already partially sepa-

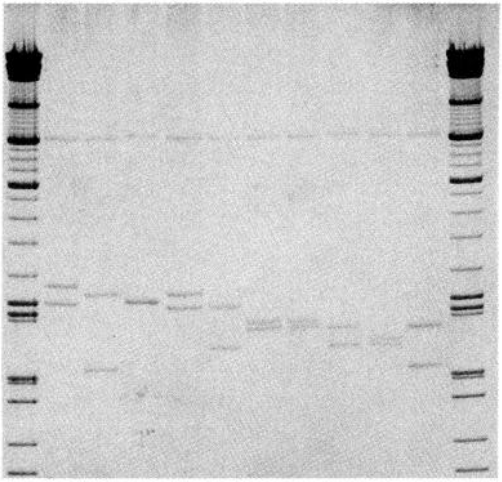


Fig. 3. Electrophoretic separation of pMCT118 (D1S80 locus). The genotypes (alleles temporarily designated), from left to right, are 13-11, 12-1, 11-11, 12-10, 10-4, 8-7, 8-7, 7-4, 5-4, and 7-1. The size of the VNTR repeat is 16 bp. The cathode is at the top.

rated sample components. As the boundary passes through each component zone, the back of the zone moves faster than the front until the zone is riding on the boundary (concentrating and sharpening the zone). Each sharpened component then rides momentarily on the boundary. The sharpened zone then unstacks as the boundary moves forward, and separation continues in a manner similar to continuous zone electrophoresis. Since this system runs at a continuous pH, but with a discontinuous voltage gradient, a wide system of operating conditions, pHs, ionic strengths, viscosities, and buffers may be employed to achieve an overall combination of variables that will achieve maximal resolution of a complex mixture of macromolecules.

6. The electrode reservoir buffer (Tris-borate) is contained in agarose (or acrylamide) reservoir plugs instead of being a free liquid contained in a buffer tray. Buffers contained in a tray or tank require paper wicks or some other bridge from the tank to the gel. A junction potential can occur between dissimilar materials (i.e., paper and polyacrylamide), thus affecting the electrophoretic process. When using agarose or acrylamide plugs (containing the trailing and counter ions) in direct contact with the resolving gel, problems with junction potential are not apparent.
7. With the system described here, resolution is 1–10 bp for fragments up to 1000 bp (Fig. 3) However, resolution potential can be manipulated by

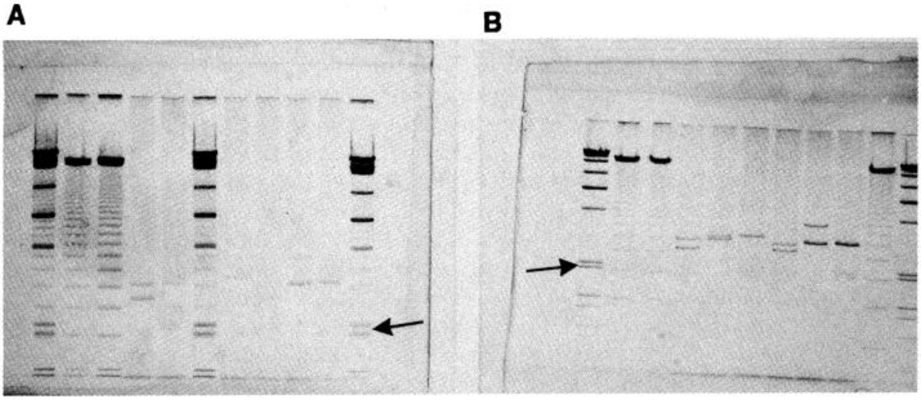


Fig. 4. Electrophoretic separation demonstrating the effect of altering the viscosity of the gel. The alleles of the hypervariable region of the apolipoprotein B gene for the same samples are clearly separated further in a gel with no glycerol **(A)** compared with a gel containing 7.1% glycerol **(B)**. The arrows indicate the increased separation of the 506 and 517-bp fragments of the 1-kb ladder (BRL, Gaithersburg, MD) as well. The cathode is at the top.

changing either the molarity of the Tris-sulfate in the gel or the viscosity (i.e., glycerol concentration) of the gel (Fig. 4). The decision to employ any particular approach depends on the size of DNA fragments and their resolution requirements. The formulations provided in this chapter are given as guidelines only.

Evaluation is underway of alternative discontinuous buffer systems with different moving boundary characteristics, such as using borate or glycine as trailing ions and citrate, chloride, or formate as leading ions.

Acknowledgments

This is publication number 90-02 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only and inclusion does not imply endorsement by the Federal Bureau of Investigation.

References

1. Wyman, A. R. and White, R. (1980) A high polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. USA* **77,** 6754–6758.
2. Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. (1987) Variable number of tandem repeat markers for human gene mapping. *Science* **235,** 1616–1622.
3. Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98,** 503–517.

4. Budowle, B. and Baechtel, F S (1990) Modifications to improve the effectiveness of restriction fragment length polymorphism typing. *Appl. Theor. Electro.* **1**, 181–187.
5. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354.
6. Saiki, R. K., Bugawan, T. L, Horn, G. T., Mullis, K. B., and Erlich, H. A. (1986) Analysis of enzymatically amplified betaglobin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes *Nature*, **324**, 163–166.
7. Horn, G. T., Richards, B., and Klinger, K. W. (1989) Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. *Nucleic Acids Res.* **17**, 2140.
8. Kasai, K., Nakamura, Y., and White, R. (1989) Amplification of a VNTR locus by the polymerase chain reaction (PCR), in *Proceedings of an International Symposium on the Forensic Aspects of DNA Analysis* (Government Printing Office, Washington, DC), in press
9. Boerwinkle, E., Xiong, W., Fourest, E., and Chan, L. (1989) Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction. Application to the apolipoprotein B 3' hypervariable region. *Proc. Natl. Acad. Sci. USA* **86**, 212–216.
10 Allen, R. C., Graves, G., and Budowle, B. (1989) Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels and stained with silver *BioTechniques* **7**, 736–744
11. Merril, C , Goldman, D , Sedman, S., and Ebert, N. (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* **211**, 1437,1438.
12. Allen, R C , Budowle, B., Lack, P. M., and Graves, G. (1986) Rehydrated polyacrylamide gels: A comparison with conventionally cast gels, in *Electrophoresis 1986* (Dunn, M J., ed.), VCH, Weinheim, Germany, pp. 462–473.
13. Allen, R. C , Budowle, B , Saravis, C A , and Lack, P. M. (1986) Enzyme and antibody detection following isoelectric focusing on ultrathin-layer rehydrated polyacrylamide gels. *Acta Histochem. Cytochem.* **19**, 637–645.
14. Allen, R. C. (1980) Rapid isoelectric focusing and detection of nanogram amounts of proteins from body tissues and fluids. *Electrophoresis* **1**, 32–37.
15. Radola, B. J. (1980) Ultrathin-layer isoelectric focusing in 50–100 μ*M* polyacrylamide gels on silanized glass plates or polyester films. *Electrophoresis* **1**, 43–56.
16. Budowle, B. and Murch, R. S. (1987) Applications of isoelectric focusing in forensic serology, in *ACS Symposium Series*, No. 335. (Jorgenson, J. W. and Phillips, M., eds.) American Chemical Society, Washington, DC, pp. 143–157.
17. Hochstrasser, D., Patchornik, A , and Merril, C. (1988) Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining *Anal Biochem.* **173**, 412–423
18 Budowle, B. and Monson, K. L. (1989) A statistical approach for VNTR analysis, in *Proceedings of an International Symposium on the Forensic Aspects of DNA Analysis* Government Printing Office, Washington, DC, in press.
19 Allen, R. C , Moore, D J., and Dilworth, R H. (1969) A new rapid electrophoresis procedure employing pulsed power in gradient gels at a continuous pH: The effect of various discontinuous buffer systems on esterase zymograms. *J. Histochem. Cytochem.* **17**, 189,190.
20. Hjerten, S , Jerstedt, S., and Tiselius, A. (1965) Some aspects of the use of "continuous" and "discontinuous" buffer systems in polyacrylamide gel electrophoresis. *Anal. Biochem* **11**, 219–223.

CHAPTER 13

Extraction and Enzymatic Amplification of DNA from Paraffin-Embedded Specimens

Colin S. Cooper and Michael R. Stratton

1. Introduction

Histopathology archive material constitutes an enormous resource of diseased tissues. It is composed of specimens that have usually been fixed with formalin to stop further tissue changes after removal from the body and subsequently embedded in a supporting material, such as paraffin, allowing sections to be cut for examination by microscopy. It has now been shown that DNA extracted from formalin-fixed, paraffin-embedded specimens, although degraded to some extent, may be used for molecular analysis *(1–3)*. Factors determining the size range of the DNA include the length of time between removal of tissue from the patient and immersion in fixative, the nuclease content of the tissue, and the pH of the formalin. The lower the pH the more fragmented the DNA, probably owing to depurination. Unfortunately, many histopathology departments routinely use a 10% (v/v) solution of formaldehyde in water without any buffering agent. Hence, over a period of time, the formaldehyde oxidizes to formic acid and the pH drops, so that the older the formaldehyde solution the lower the pH. The length of time the specimen has spent in formalin before processing will also influence the state of the DNA. Fortunately, most small biopsy specimens will have spent an overnight period or even less in fixative.

Nucleic acids prepared from paraffin blocks can be used for enzymatic amplification of specific regions of DNA *(3)* and, when the extracted DNA is

From. *Methods in Molecular Biology, Vol. 9· Protocols in Human Molecular Genetics*
Edited by· C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

of sufficiently high mol wt, for Southern analyses *(2)*. DNA prepared from paraffin blocks can also be cloned directly into plasmid vectors *(1)*.

Enzymatic amplification by the polymerase chain reaction (PCR) results in a high level of amplification (up to 100,000-fold) of a small region of DNA (usually 100–500 bp) between two primer sequences. The amplified DNA may be sequenced (Chapter 3) or examined for mutations using a variety of methods, such as oligonucleotide hybridization (Chapters 4 and 24), denaturing-gradient gel electrophoresis (Chapter 10), and RNase mismatch cleavage (Chapter 11 and ref. *4*). Using these methods, it is possible to detect genetic alterations in tissues that have been preserved in paraffin blocks for several decades.

2. Materials

1. DNA extraction buffer: 100 m*M* Tris-HCl (pH 7.5), 100 m*M* NaCl, 10 m*M* EDTA.
2. TE: 10 m*M* Tris-HCl (pH 7.5), 1 m*M* EDTA.
3. Fluorescence measurements were made using an Aminco-Bowman spectrofluorimeter.
4. Fluorimetry buffer: 2*M* NaCl, 50 m*M* sodium phosphate, pH 7.4.
5. Hoechst 33258: A stock solution (1 mg/mL) of this reagent is stored in the dark.
6. 10× *Taq* salts: 500 m*M* KCl, 100 m*M* Tris-HCl, pH 8.3, 15 m*M* $MgCl_2$, 0.1% (w/v) gelatine. Stored at –20°C.
7. *Taq* polymerase (Cetus Corporation) is stored at –20°C and diluted to a final concentration of 1 U/µL in 1× *Taq* salts immediately prior to use.
8. dNTP mixture: A 40-m*M* stock solution, containing 10 m*M* dATP, 10 m*M* dGTP, 10 m*M* dCTP, and 10 m*M* dTTP is stored at –20°C.
9. PCR oligonucleotide primers (usually 20 mers) are dissolved at 40 µg/mL ($OD^{1\,cm}_{260\,nm}$ = 1.0) and stored at –20°C.
10. Stock solutions of 3*M* sodium acetate, 10% (w/v) SDS, 95% ethanol, 70% ethanol and chloroform:isoamylalcohol (24:1, v/v) should be prepared, and can be stored at room temperature. Stock solutions of proteinase K (25 mg/mL) should be stored at –20°C. Phenol:chloroform (1:1, v/v) is prepared fresh as required. Phenol is buffered and water saturated prior to use by the addition of an equal vol of 100 m*M* Tris-HCl (pH 7.5).

3. Methods

3.1. Extraction of DNA

1. Cut 25 20-µm sections from the paraffin block. Cutting to a total depth of 0.5 mm of tissue is acceptable to most pathologists (*see* Note 1).

2. Place the sections in a 15-mL polypropylene screw top tube and add (i) 1 mL of DNA extraction buffer, (ii) 100 µL 10% (w/v) SDS, and (iii) 20 µL of an aqueous solution of proteinase K (25 mg/mL). Mix by vortexing and incubate for 16–24 h at 48°C (*see* Note 2). Add a further 0.5 mL of extraction buffer, 50 µL of 10% (w/v) SDS, and 10 µL of proteinase K solution (25 mg/mL). Mix by vortexing and incubate for additional 1–6 d at 48°C (*see* Note 3).
3. The mixture should now consist of a discolored fluid overlaid with a slurry of paraffin fragments. Add 1.5 mL of phenol and mix by rotation or gentle shaking for 20 min (*see* Note 4). After the mixture has been centrifuged at 4000 rpm for 10 min in a bench-top centrifuge, the aqueous phase is collected (*see* Note 5).
4. Divide the crude aqueous phase into three 1.5-mL Eppendorf tubes (0.5 mL per tube) and subject the mixture present in each tube to a further extraction with an equal vol (0.5 mL) of phenol. Centrifuge the mixture in a microcentrifuge for 5 min and save the aqueous phase. Repeat this extraction step using an equal vol of (i) phenol:chloroform (1:1; v/v) and (ii) chloroform:isoamylalcohol (24:1, v/v).
5. Add 1/10 vol of 3*M* sodium acetate to the aqueous phase. Then add 2 vol of ethanol and allow the nucleic acid to precipitate at –20°C for 16 h (*see* Note 6). Pellet the precipitated nucleic acids by centrifugation for 10 min in a microcentrifuge. The pellet of nucleic acids should then be washed with 70% ethanol, collected by centrifugation, air-dried, and dissolved in 100 µL TE.
6. The integrity of the nucleic acids can be monitored by electrophoresis in 1–2% (w/v) agarose gels (Fig. 1).

3.2. DNA Concentration

The protocol described above yields a mixture of DNA and RNA (Note 7). The standard diphenylamine assay can be used to determine the concentration of DNA in this mixture. However, because the DNA is often at low concentration, we prefer to use a highly sensitive fluorescence assay that requires only very small amounts of nucleic acids.

1. Fluorescence measurements are made using an emission wavelength of 385 nm and an excitation wavelength of 290 nm. The fluorescence scale is set to zero using a quartz cuvet containing 1 mL of fluorimetry buffer, to which 1 µL of a stock solution (1 mg/mL) of Hoechst 33258 dye has been added.
2. The machine is calibrated by adding 1 µL aliquots of a solution of DNA of known concentration (usually 0.1–1 µg/mL) and noting the increase in fluorescence.

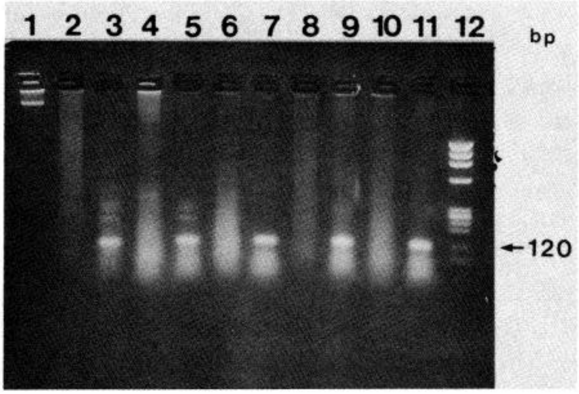


Fig. 1. Amplification by the PCR of DNA from paraffin blocks. DNAs were prepared from a cell line (lane 1) and formalin-fixed paraffin-embedded tissue (lanes 2,4,6,8, and 10). Aliquots of the DNA shown in lanes 2,4,6,8, and 10 were subjected to 40 cycles of PCR using primers that spanned codon 12 of the H-*ras* allele and analyzed in, respectively, lanes 3,5,7,9, and 11. Lane 12 contains DNA size markers. DNA was electrophoresed in a 2% (w/v) agarose gel, stained with ethidium bromide (0.1 µg/mL in electrophoresis buffer), and visualized over UV light. The bright band at 120 bp represents the product of the PCR.

3. Add 1 µL of test solutions to the same cuvet and calculate the increase in fluorescence over the previous reading. Aliquots (1 µL) of several test solutions can be examined in a similar way before the Hoechst 33258 reagent needs to be changed.
4. The concentration of the extracted DNA can be calculated using the following equation:

$$\frac{\text{Concentration of DNA in test solution}}{\text{Concentration of calibration DNA}} = \frac{\text{Fluorescence increase observed for test DNA}}{\text{Fluorescence increase observed for calibration DNA}}$$

For use in the PCR, DNAs should be diluted to 0.1 mg/mL using TE.

3.3. Polymerase Chain Reaction (PCR) (Method 1)

One microgram of each archival sample is routinely used for enzymatic amplification. However, if the yields are low, a smaller amount of DNA can be used. If the PCR primers are between 100 and 200 bp apart, efficient amplification of specific DNA sequences can be achieved with even the most degraded DNA samples (*see* Note 8). For a more detailed description of the PCR procedure, *see* Chapters 1 and 5.

1. The PCR incubation mixture contains 55 µL H_2O, 10 µL 10× *Taq* salts, 2 µL of dNTP mixture (containing 10 m*M* dATP, 10 m*M* dCTP, 10 m*M* dTTP, and 10 m*M* dCTP), 10 µL of oligonucleotide primer A, 10 µL of oligonucleotide primer B, and 10 µL of a 0.1 µg/mL solution of DNA. Finally, 3 U of *Taq* polymerase (3 µL) are added and the mixture is overlayed with liquid paraffin to prevent evaporation and subjected to 40 PCR cycles; each cycle involves denaturation at 94°C for 2 min, primer annealing at 55°C for 2 min, and chain elongation at 72°C for 2 min.
2. The success of the amplification can be monitored by subjecting aliquots of the amplification mixture to electrophoresis in 2% (w/v) agarose gels (Fig. 1). The amplified DNA segments can be visualized by staining with ethidium bromide (0.1 µg/mL).

3.4. Polymerase Chain Reaction (PCR) (Method 2)

Sometimes it is more convenient to perform the PCR directly on tissue fragments rather than on the DNA extracted from them *(4,5)*. This is a particularly useful approach when only a small tissue fragment has been embedded.

1. Cut 20-µm sections from the paraffin block as described previously. Place one or two sections in a 0.5-mL Eppendorf tube and add 400 µL of xylene to deparaffinize the section. Vortex briefly and collect the tissue fragments by centrifugation in a microcentrifuge for 5 min (*see* Note 9).
2. The tissue fragments should then be washed with 95% ethanol (400 µL), collected by centrifugation, and air-dried.
3. Add the ingredients of the PCR mixture described above directly to the deparaffinized tissue fragment. Incubate the mixture at 94°C for 10 min, add 3 U of *Taq* polymerase, overlay with 100 µL liquid paraffin, and commence PCR cycling as described above.

4. Notes

1. The efficiency of DNA extraction may be improved by cutting thinner sections, but this renders the sections much more difficult to handle. We store sections at room temperature in universal bottles (20-mL size) until required and do not usually remove the excess paraffin from around the tissue sample. The paraffin present in these sections may be removed by xylene extraction (*see* Polymerase chain reaction, Method 2). However, in our hands, this extraction step does not increase the efficiency of DNA extraction or alter the size distribution of the extracted DNA.

2. Paraffin sections that adhere to the side of the tube may be returned to the main extraction mixture by vortexing the mixture after the incubation has been allowed to proceed for 2–3 h. It should be noted that the lower mol wt DNA fragments emerge first from the tissue. Thus, if only high mol wt fractions are required, the tissue fragments should be collected by centrifugation after the initial 16–25 h incubation and resuspended in fresh extraction buffer, SDS, and proteinase K.
3. The length of time that should be allowed for this incubation step depends on the nature of the tissue. We routinely allow the incubation to proceed for 48 h, but for tissues with large amounts of dense collagenous stroma, for example, uterine muscle, a longer incubation of up to 5 or 6 d may increase the yield of DNA.
4. Vortexing may be necessary to dislodge the paraffin plug, which becomes solid when maintained at room temperature, and which may hinder the mixing of phenol with the aqueous phase.
5. To maximize the yield of DNA, it is advisable at this stage to collect as much of the aqueous phase as possible. Don't worry if some of the paraffin that sits at the interface and some of the phenol layer is also collected. This material can be removed in the subsequent round of phenol extraction.
6. If DNA of sufficiently high mol wt is present, it can, after precipitation with ethanol, be spooled onto a glass rod, washed with 70% ethanol, and directly redissolved in TE.
7. The mixture of nucleic acids can be used directly for enzymatic amplification. However, if required, the RNA can be removed by digestion with DNase-free pancreatic RNase (50 µg/mL in extraction buffer for 30 min at 37°C). The mixture is then treated with proteinase K (0.2 mg/mL in extraction buffer containing 0.5% (w/v) SDS for 60 min at 37°C) and protein is removed by sequential extraction with equal vol of (i) phenol, (ii) phenol-chloroform (1:1, v/v), and (iii) chloroform:isoamylalcohol (24:1, v/v). Finally, add 1/10 vol of 3*M* sodium acetate and precipitate the DNA by adding 2 vol of ethanol and incubating the mixture for 16 h at –20°C.
8. As the DNA becomes more degraded, an increasing number of DNA fragments will be broken between the PCR primer sequences and will, therefore, not function as template during enzymatic amplification. If the positions of the breaks in the DNA are distributed randomly, the number of unbroken sequences that are available for amplification by the PCR is determined by the Poisson distribution and can be calculated using the equation:

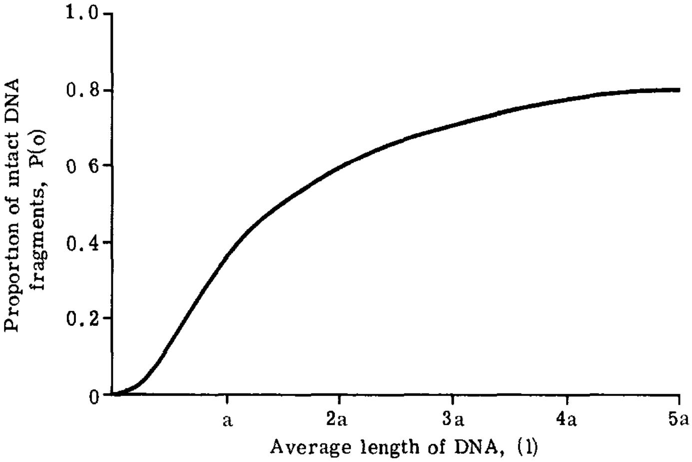


Fig. 2. Effect of DNA degradation on the PCR. The proportion of unbroken DNA sequences that are available for the PCR reaction [P(o)] depends on the average size of the DNA *(1)*, and can be calculated from the equation $P(o) = e^{-a/l}$, where a is the distance between the PCR primers.

$$P(o) = e^{-a/l}$$

where P(o) is the proportion of fragments without a break between the two PCR primer sequences, a is the distance between the primers, and l is the average length of the DNA. For example, in the hypothetical example where the DNA has been degraded to an average length of only 200 bp and the distance between the primers is 200 bp, 37% of DNA fragments will remain unbroken between the primer sequences (Fig. 2). In practice, this means that efficient amplification can usually be achieved for even the most degraded DNA samples.

9. Sometimes the tissue fragment does not pellet at the bottom of the tube. In such a case, remove as much xylene as possible with a Pasteur pipet and proceed to the next step.

Acknowledgments

CSC is funded by the Cancer Research Campaign, UK.

References

1. Goelz, S E , Hamilton, S R , and Vogelstein, B. (1985) Purification of DNA from formaldehyde fixed and paraffin embedded tumour tissue *Biochem Biophys Res Comm* **130**, 118–126

2. Dubeau, L., Chandler, L. A., Gralow, J R., Nichols, P. W., and Jones, P. A. (1986) Southern blot analysis of DNA extracted from formalin-fixed pathology specimens *Cancer Res* **46,** 2964–2969.
3. Imprain, C. C., Saiki, R. K., Erlick, A. A., and Templitz, R. L. (1987) Analysis of DNA extracted from formalin-fixed paraffin embedded tissues by enzymatic amplification and hybridization with sequence specific probes. *Biochem. Biophys. Res. Comm.* **142,** 710–716.
4 Almoguera, C., Shibata, D., Forrester, K., Martin, J , Arnheim, N , and Perucho, M. (1988) Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell* **53,** 549–554.
5. Shibata, D., Arnheim, N., and Marta, J. (1988) Detection of human papilloma virus in paraffin embedded tissue using the polymerase chain reaction. *J. Exp. Med.* **167,** 225–230

CHAPTER 14

The Use of the Polymerase Chain Reaction in the Mapping of Human Genes Using Somatic Cell Hybrids

Cathy Abbott and Sue Povey

1. Introduction

The majority of human–rodent somatic cell hybrids lose human chromosomes in a more or less random manner. It is possible to obtain a panel of hybrids in which each contains a unique subset of human chromosomes. In principle, human genes can be mapped in such hybrids by looking for correlation between the presence of a particular protein product or DNA sequence and a particular human chromosome. Hybrids containing well defined fragments of human chromosomes can then be used for regional localization. The availability of DNA from well characterized hybrids allows the rapid assignment of a newly cloned human sequence without the expertise and expensive equipment currently required for analysis of direct *in situ* hybridization to metaphase spreads or flow sorting of human chromosomes.

Somatic cell hybrids are relatively unstable in culture, thus each time cells are grown for mapping studies, their human chromosome content has to be reassessed. This can be done by direct karyotype analysis or by looking for gene products or DNA sequences known to lie on particular chromosomes. Direct karyotype analysis requires considerable time and expertise, and traditionally, many hybrids have been characterized by isozyme analysis that separates human and rodent gene products, or by detection of surface antigens coded by human genes. Increasingly, hybrids are characterized us-

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

Table 1
Sequences of Oligonucleotide Primers Used for Detecting the Presence of Some Human Chromosomes in Somatic Cell Hybrids

Chromosome	Gene	Primers used	Ref.
3	Unpublished	5' GATTGGATCTCTTCCTTTTGATGAC 3' 5' CTGGGGGAGGAAAACTCAATAAAAT 3'	
5	C9	5' TAGATACATTGAGTCTCTCCTGATT 3' 5' CAGTCTATCACAATGAGAGAGATGG 3'	*1*
7	PGAM2	5' GGTCCTAGACTCAACTCCGTGCCAC 3' 5' TCTGGCCTTGTGGAAGGTACCAGGC 3'	2
9	Aldolase B	5' TCATTGCTTGCTTTCTCAAGCAGGG 3' 5' CAATGCTTCTCCGTGTTGGAAAGTC 3'	*3*
14	α1 Antitrypsin	5' CTGGTGATGCCCACCTTCCCCTCTC 3' 5' GTCACCCTCAGGTTGGGGAATCACC 3'	*4*
17	Unpublished	5' GAAGATGGGAAGTCCTGTTTTGCCC 3' 5' AGCAGATGGTTAGGGTACTAGTGGG 3'	*5*

ing Southern blots of hybrid DNA hybridized with probes, which distinguish between human and rodent genes. More recently, species-specific polymerase chain reaction (PCR) has been used to assist in the characterization of hybrids and in the assignment of several new genes *(1–3)*. For the characterization of hybrids, oligonucleotide primers are synthesized to correspond to (usually) intron sequences of a human gene that has been mapped to a particular chromosome, and for which genomic sequence is available. Intron sequences, in our experience, have always been sufficiently divergent between human and rodents that only the human gene (and not its rodent homolog) will be amplified by PCR. Thus, a band is seen only in those hybrids in which the corresponding human chromosome is present. Table 1 shows some of the primers routinely used in our laboratory for this purpose. For the mapping of a new gene, a similar strategy may be employed. Below we outline some of the practical details of this approach.

1.1. Strategy

1.1.1. Choice of Oligonucleotides

In the ideal case, sequence data would be available for the human gene and its homologs in mouse, rat, and hamster. However, this is seldom the

case. In our experience, primers synthesized to correspond to human intron sequences have proved to be perfectly adequate. We have found 25 mers to be ideal. The ideal distance apart for the primers seems to be 200–500 bp. If only human cDNA sequence is available, then it is probably safer to map the gene using conventional Southern blotting to avoid the potential problem of the sequence to be amplified having a large (unsuspected) intron in it. If blots are unsatisfactory for any reason, then primers from the 3' untranslated region of the cDNA are the best bet. If rodent cDNA sequence is available for comparison, then try to maximize mismatches between human and rodent sequence. Mismatches should be at the 3' ends of the primers to minimize coamplification of the rodent gene.

When both rodent and human genomic sequences are available, then it might be possible to synthesize oligonucleotides corresponding to exon sequences, which will amplify both human and rodent genes to give products of different lengths. This will mean that DNA preparations that cannot be amplified can be immediately identified.

1.1.2. Choice of Conditions

In general, it is the annealing step of the PCR that is critical in species-specific PCR. The annealing temperature should be as high as possible for the human gene to be amplified, and annealing time should be kept short. As a rule of thumb, 30 s at 55°C is usually ideal for 25 mers. For all the primer pairs given in Table 1, we have found 30 cycles, each comprising 30 s at 90°C, 30 s at 55°C, and 30 s at 70°C, to be ideal. Clearly, when primers have first been prepared, they must be checked on all human and rodent parent DNAs before proceeding to hybrid DNAs, preferably starting with an excess of the rodent parent to ensure specificity. In all hybrids for which a negative result is obtained, it is essential to check that the DNA is capable of being amplified. This is most easily done with control primers that will amplify a particular sequence in any rodent DNA, as the human component of the hybrid will obviously vary. It may be necessary to use more DNA from the hybrid than would be used for a human control, since a chromosome will not necessarily be present in all cells of a hybrid. If the hybrids have been characterized by PCR initially, then the mapping PCR should be directly comparable to the characterization. However, in general, we use 2–3 µg hybrid DNA compared with 1 µg human DNA.

2. Materials

1. A programmable machine capable of cycling temperatures or three water baths.
2. Dimethyl Sulfoxide (DMSO). This can be bought as a liquid and stored at room temperature.

3. 10× *Taq* polymerase buffer: 166 µ*M* $(NH_4)_2SO_4$, 0.67*M* Tris-HCl, pH 8.8, 67 µ*M* $MgCl_2$, 100 µ*M* 2-mercaptoethanol, 67 µ*M* EDTA, and 1.7 mg/mL bovine serum albumin (BSA), or as supplied by manufacturer. Store at –20°C
4. dATP, dCTP, dGTP, and dTTP: 15 m*M* Each. These can most cheaply be purchased as powders, which should be made up in distilled water and brought to pH 7 with unbuffered Tris or NaOH. Store at –20°C.
5. Oligonucleotide primers: Store at –20°C.
6. *Taq* polymerase (from, for example, Anglian Biotechnology or Perkin Elmer Cetus). Store at –20°C.
7. Light liquid paraffin. Store at room temperature.
8. Agarose: Low EEO from Sigma, or Nusieve agarose from FMC Bioproducts if small products are to be separated.

3. Method

1. Mix the following together in a 1.5-mL Eppendorf tube:
   a. 10 µL of 10× PCR buffer.
   b. 10 µL of DMSO.
   c. 10 µL each of 15 m*M* dATP, dCTP, dGTP, and dTTP.
   d. 50 pmol each oligonucleotide.
   e. 1–3 µg DNA.
   f. H_2O to 100 µL.
2. Heat at 95°C for 5 min.
3. Remove the tubes from the heating block or water bath, add 2–3 U *Taq* polymerase to each tube, and layer 100 µL of light paraffin oil on top. Replace immediately in machine or first water bath.
4. Carry out 30 cycles PCR under chosen conditions (*see* Section 1.1).
5. When cycles are complete, pipet the paraffin off the tube. Remove a 10-µL aliquot from the tube and analyze by running an agarose minigel (at, for example, 2%) and visualize by ethidium bromide staining under UV light.

An Example of typical PCR is shown in Fig. 1.

4. Notes

1. For the assignment of a human gene to a chromosome, a panel of between 10 and 20 hybrids, preferably independently derived, is required. There is one commercial source of amplifiable DNA from somatic cell hybrids, although we have no personal experience of this. The supplier

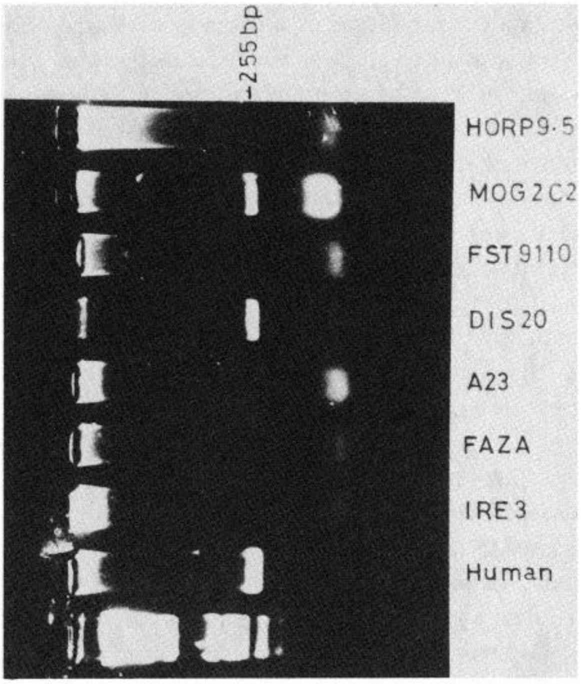


Fig. 1. PCR of complement component C9 in some somatic cell hybrids and their parent cell lines.

is Bios Corporation, New Haven, CT. Alternatively, one of us (S. P.) is able to suggest people with somatic cell hybrid DNA who might be willing to collaborate. For a confident exclusion of any given chromosome as the site of a gene, it is desirable to have at least three examples of discordance between gene and chromosome. This should include one example in each direction, i.e., absence of gene with presence of chromosome and vice versa. If this is not adhered to, occasional mistakes will be made because the sensitivity of detection of the marker and of the newly mapped gene may not be the same, i.e., PCR is generally considered to be more sensitive than Southern blotting, and so, it is formally possible that the presence of the gene being mapped might be detected in a hybrid that had been typed as being negative for the chromosome to which the gene actually maps.

2. Newly made hybrids are often more unstable than old established lines, and subcloning is usually necessary to obtain reasonably homogeneous cell populations.

3. In the initial characterization of a hybrid, it is extremely useful to have direct karyotype analysis since many hybrids contain fragments of chromosomes not detected by the markers, and the presence of a marker is not a reliable guide to the presence of a normal chromosome. Subsequent batches of the hybrids can be reasonably characterized by markers only; usually, it is sufficient to check that chromosomes previously present have been retained, but it is unfortunately not unknown for a subpopulation of cells containing a chromosome not previously detected to take over the culture.
4. It is theoretically possible to combine different pairs of primers when characterizing a hybrid in order to assess the presence of more than one chromosome in a reaction. In practice, however, we have found this to be less reliable and reproducible than setting up a different reaction for each primer pair, possibly because of the stochastic nature of the first few rounds of PCR.
5. We have found that even under standard PCR conditions, it is generally possible to make some judgment of the proportion of cells in a hybrid likely to contain a given chromosome. This can only be done, however, by direct comparison with a human control or a well-characterized hybrid amplified under identical conditions, and should not be relied on.

References

1. Abbott, C , West. L., Povey, S., Jeremiah, S., Murad, Z , Discipio, R., and Fey, G (1989) The gene for human complement component C9 mapped to chromosome 5 by polymerase chain reaction. *Genomics* 4, 606–609.
2. Edwards, Y., Saburo, S., Schon, E., and Povey, S. (1989) The gene for human muscle-specific phosphoglycerate mutase, PGAM2, mapped to chromosome 7 by PCR *Genomics* 5, 948–951.
3 Jeremiah, S., Abbott, C. M., Murad, Z., Povey, S., Thomas, H. J , Solomon, E., Discipio, R., and Fey, G H. (1990) The assignment of the genes coding for human complement components C6 and C7 to chromosome 5. *Ann. Hum. Genet.* 54, 141–147.
4. Tolan, D. R. and Penhoet, E. E. (1986) Characterization of the human aldolase B gene *Mol. Biol. Med.* 3, 245–264.
5. Abbott, C , Povey, S., Vivian, N , and Lovell-Badge, R. (1988) PCR as a rapid screening method for transgenic mice *Trends in Genet* 4, 325.

Chapter 15

The Southern Blot

An Update

Michael R. Evans, Andrew L. Bertera, and Dennis W. Harris

1. Introduction

The Southern blot technique, published by Southern in 1975 *(1)*, is probably the most widely used method in molecular biology and, to date, has received in excess of 100,000 literature citations. The purpose of the method is the detection of gel-fractionated DNA molecules following transfer to a membrane. The Southern blot is used extensively in research applications (e.g., gene mapping) as well as in diagnosis (e.g., DNA fingerprinting, Chapter 22; inherited disorders, Chapter 30).

Here, we concentrate on practical improvements to the techniques of Southern blotting, probe labeling, and hybridization, introduced since the previous review in this series *(2)*. The most significant recent advances in the methodology are as follows:

1. New membranes: Supported nylon membranes are now the media of choice for the immobilization of target nucleic acids, and have largely supplanted nitrocellulose for this purpose. Nylon has improved DNA binding capacity, higher tensile strength, and rapid protocols are now available for the covalent fixation of target DNA to the membrane. Furthermore, the robust nature of nylon membranes allows many successive probings to be conducted with minimal losses in sensitivity.

From: *Methods in Molecular Biology, Vol 9. Protocols in Human Molecular Genetics*
Edited by C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

Table 1
Major Properties of Uniform Labeling Methods

	NT	RP	RNA
Template DNA	ds	ss or ds*	ds
Amount of template	0.5–1 μg	25 ng	1–2 μg
Reaction time	~2 h	0.5–3 h	~1 h
Incorporation efficiency	~60%	~75%	~75%
Nature of probe	DNA	DNA	RNA
Amount of probe	0.5–1 μg	40–50 ng	~250 ng
Potential specific activity (dpm/μg)	5×10^8	5×10^9	5×10^9
Probe length	~500 bp	~200 bp	Defined
Insert specific	No	No	Yes
Subcloning required?	No	No	Yes
Labeling in agarose	Yes	Yes	No

NT = Nick translation
RP = Random primer
RNA = RNA polymerase.
*After denaturation.

2. Transfer methods: The Southern blot can be generated by a range of transfer methods, including vacuum *(3)* or electrophoretic blotting, in addition to the original capillary method *(1)*. The prime advantage of vacuum blotting is speed: gel pretreatment and transfer can be completed in only 40 min for a wide range (<1 – >20 kb) of DNA fragment sizes. However, both the vacuum and electrophoretic blotting methods require specialized equipment, in contrast to the capillary technique, which is cheaper and easy to set up with basic laboratory materials. Moreover, optimized capillary blotting conditions can give excellent transfer in as little as 2 h *(4–6)*.
3. Probe labeling: Table 1 summarizes the features of the three major radioactive probe labeling methods. Random primer labeling *(7,8)* is currently the most popular technique because it generates high specific activity ^{32}P-labeled probes with relatively small amounts of template DNA (e.g., 25 ng) in times as short as 30 min. However, the other labeling methods are still used to a significant degree. Most protocols used for nick translation produce more probe than do other uniform labeling methods. It is therefore a particularly suitable reaction when carrying out multiple hybridizations with the same probe or when a high probe

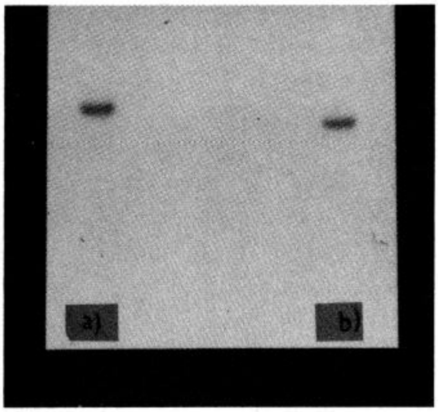


Fig. 1. A comparison of the rapid hybridization system-Multiprime with conventional overnight hybridization. *Hind* III-digested human placental DNA (5 µg) was blotted onto Hybond-*N* nylon membrane and hybridized with a ^{32}P-labeled probe. The probe was generated by random primer labeling using the Multiprime system, and was used at a concentration of 8 ng/mL in the hybridization. Hybridizations were for (a) 2 h using the rapid hybridization buffer, and (b) 16 h using a conventional hybridization solution. Autoradiography was for 4 h at –70°C with Hyperfilm-MP and two intensifying screens.

concentration is desired. RNA probes are now preferred by some researchers, and there are several reports of increased sensitivity when using these probes (e.g., *see* ref. 9). This is probably owing to the fact that RNA probes cannot reanneal, unlike double-stranded DNA probes. In addition, under certain conditions (for example, in the presence of 50% [v/v] formamide), RNA:DNA hybrids are significantly more stable than DNA:RNA hybrids.

Several methods are now available for the nonradioactive labeling and detection of RNA and DNA. Chapter 16 describes these techniques in more detail.

4. Rapid hybridization: Filter hybridizations proceed at approx 10-fold lower rates than the corresponding solution reactions *(10)*. The prolonged incubation required is probably the major limitation of the blot hybridization technique. However, recent work in our laboratory has led to the development of a rapid hybridization system that reduces the hybridization time from 16 to 2 h, without any loss in sensitivity. This system is based on the use of specialized hybridization rate enhancers that are effective, even at low probe concentrations (Fig. 1).

2. Materials

1. Restriction endonuclease buffer; prepare a 10× stock, according to the manufacturer's instructions. Several suppliers provide buffer concentrates together with the restriction enzyme.
2. Agarose gels: Use a low electroendosmosis grade agarose, e.g., Sigma A-6031.
3. Electrophoresis buffer: Prepare a 50× stock of Tris-acetate-EDTA buffer (TAE) consisting of 2*M* Tris base, 0.05*M* disodium EDTA, adjusted to pH 8 using glacial acetic acid.
4. Loading buffer: 0.05% (w/v) Bromophenol blue, 0.05% (w/v) xylene cyanol, 50% (v/v) glycerol, and 0.05*M* EDTA in TAE buffer.
5. Gel pretreatment solutions:
   a. Depurination solution: 0.25*M* HCl.
   b. Denaturation solution: 1.5*M* NaCl, 0.5*M* NaOH.
   c. Neutralization solution: 1.5*M* NaCl, 0.5*M* Tris-HCl, pH 7.5.
6. Transfer buffers: High salt transfer (20× SSC) 3*M* NaCl, 0.3*M* trisodium citrate, pH 7.
   Alkaline transfer: 0.4*M* NaOH.
7. Transfer apparatus: Capillary transfer can be carried out in a glass or perspex tray (a typical setup is described in ref. *2*). Vacuum and electrophoretic blotting equipment is available commercially.
8. Blotting membranes: Neutral or positively charged nylon membranes are available from various suppliers; Amersham products are Hybond-N and Hybond-N+, respectively.
9. DNA fixation: A transilluminator of maximal output wavelength approx 312 nm is required for crosslinking DNA to neutral nylon. If Hybond-N+ is used, prepare the following solutions:
   a. Fixation solution: 0.4*M* NaOH.
   b. Rinse solution: 5× SSC (1 in 4 dilution of solution 6).
10. Probe labeling: The preferred method is random primer labeling (*see* Introduction). DNA labeling kits are available for this purpose (e.g., Amersham Multiprime codes RPN1600 and 1601), as are a complete range of ^{32}P- and ^{35}S-labeled nucleotides.

    The DNA labeling system consists of the following components required for probe labeling:
    a. Solution 1: dATP, dGTP, and dTTP in a buffer containing 250 m*M* Tris-HCl, pH 7.8, 25 m*M* magnesium chloride, and 50 m*M* 2-mercaptoethanol.
    b. Solution 2: 1.8 mg/mL Random hexanucleotides in an aque-

ous solution containing nuclease-free bovine serum albumin (BSA) at 4 mg/mL.

c. Solution 3: 1 U/μL of cloned DNA polymerase "Klenow" fragment in 50 m*M* potassium phosphate, pH 6.5, 10 m*M* 2-mercaptoethanol, and 50% (v/v) glycerol (*see* refs. *7,8* for further details of probe labeling by the random primer technique).

11. Hybridization buffer: 5x SSC, 5x Denhardt's solution (prepare as a 100x stock solution), 0.5% (w/v) SDS. This buffer is suitable for conventional overnight hybridizations.
    a. 2% (w/v) BSA.
    b. 2% (w/v) Ficoll.
    c. 2% (w/v) PVP (polyvinylpyrollidone).

    Alternative hybridization buffers are available that allow hybridization time to be reduced from 16 h to as little as 2 h without any reduction in detection sensitivity (*see* Note 12).
12. Stringency washes:
    a. Wash 1: 2x SSC, 0.1% (w/v) SDS.
    b. Wash 2: 1x SSC, 0.1% (w/v) SDS.
    c. Wash 3: 0.1–0.7x SSC, 0.1% (w/v) SDS.
13. Hybridization container: Hybridization can be performed in dedicated hybridization boxes *(2)* or in heat-sealable plastic bags and the incubation performed in a shaking water bath. Alternatively, roller bottles can be used in "rotisserie" ovens, which are now commercially available.
14. Autoradiography: Use a suitable X-ray film (e.g., Hyperfilm-MP, Amersham) together with two intensifying screens.

3. Methods

3.1. Restriction Endonuclease Digestion

1. Prepare the sample DNA using the appropriate method (e.g., *see* Volume 2, this series, or ref. *11* for a range of methods suitable for genomic DNA, plasmid DNA, or phage DNA).
2. Digest the DNA sample with a restriction enzyme according to the manufacturer's instructions, using at least 2 U of enzyme/μg of DNA. Complete digestion of genomic DNA usually requires prolonged incubations (e.g., 16 h).
3. At the end of the reaction, place the samples on ice. Remove a small aliquot and check on an agarose minigel for complete digestion before proceeding with the Southern blot gel.

3.2. Agarose Gel Electrophoresis

1. Make a 0.6–2% (w/v) agarose gel by adding agarose powder to electrophoresis buffer and heating to approx 90°C. Cool the molten agarose to 50–60°C, add ethidium bromide to a final concentration of 1 µg/mL, and pour into a gel former. Insert the gel comb to produce wells of up to 5 mm in width (*see* Note 1).
2. Allow the gel to set and then place in an electrophoresis tank. Fill the tank with electrophoresis buffer, sufficient to cover the surface of the gel.
3. Add 0.1 vol of the gel loading buffer to the DNA samples. For higher eukaryotic genomic DNA, load 1–10 µg of the sample. Lower amounts (e.g., nanogram quantities) are required for less complex DNAs, such as plasmid or phage. A mol wt marker sample should also be included on the gel (*see* Note 2).
4. Electrophorese at constant voltage (e.g., 60 V over 4–5 h for a 20-cm long gel) and run until the bromophenol blue dye is at least two-thirds of the way down the gel (*see* Note 3).

3.3. Gel Pretreatment

1. After electrophoresis, place the gel in depurination solution and agitate slowly on an orbital shaker. Leave until the dyes have changed color (*see* Note 4) plus a further 10 min. If an alkaline transfer is to be performed, proceed directly to the blotting step (*see below*).

   For high salt transfers, continue the gel pretreatment as follows:
2. Rinse the gel in distilled water and place in denaturation buffer. Leave for 30 min with shaking.
3. Rinse the gel in distilled water and place in neutralization buffer. Agitate for a further 15 min. Repeat once.

3.4. Southern Blotting

Electrophoretic/vacuum blotting: Follow the manufacturer's instructions for efficient transfer.

Capillary blotting:

1. Fill a tray or glass dish with 20× SSC. Make a platform and cover it with a wick made from three sheets of Whatman 3MM filter paper, saturated with 20× SSC. If an alkaline transfer is to be performed, substitute the SSC with 0.4*M* NaOH in all stages of the capillary blotting process.
2. Place the gel on the wick and avoid trapping air bubbles beneath it. Surround with cling film to prevent the blotting buffer being absorbed directly into the paper towels above.

3. Cut a sheet of nylon membrane to the exact size of the gel and place on top of the gel. Avoid trapping air bubbles under the membrane (*see* Note 5).
4. Place three sheets of 3MM paper, cut to size and wetted with 20× SSC, on top of the nylon membrane.
5. Place a stack of absorbent paper on top of the 3MM paper.
6. Place a glass plate on top of the absorbent towels and put a 0.75 kg weight on top. Allow DNA transfer to proceed for 2–16 h.
7. After blotting, carefully dismantle the apparatus. Before removing the gel, mark the membrane with a pencil to allow later identification of the individual tracks.

3.5. Fixation of DNA to the Membrane

3.5.1. Neutral Nylon/High-Salt Transfers

1. Allow the membrane to air-dry for up to 1 h at room temperature or 10 min at 80°C.
2. Wrap the membrane in Saran Wrap™ (Dow Chemical Co.) and place DNA side down on a UV transilluminator for 2–5 min. The precise exposure time should be determined by prior calibration of the transilluminator (*see* Note 6).

3.5.2. Hybond-N+/High Salt Transfers

1. Place the membrane on a pad of three sheets of 3MM paper soaked in 0.4*M* NaOH. Leave for between 2 and 60 min at room temperature. This treatment efficiently fixes the target DNA to the Hybond-N+ membrane.
2. Rinse the membrane briefly in 5× SSC with gentle agitation (maximum time, 1 min).

3.5.3. Hybond-N+/Alkaline Transfers

The alkaline transfer method *(5)* results in crosslinking of the target DNA to the membrane during transfer. There is therefore no need for a posttransfer fixation step after alkaline transfer to positively charged nylon membranes.

After fixation by any of these methods, the membrane can be used directly in the prehybridization step or wrapped in Saran Wrap™ and stored at 4°C.

3.6. Probe Labeling

Random primer labeling protocols typically generate probes of a specific activity ~2×10^9 dpm/µg of template DNA.

1. Dilute the DNA to be labeled to a concentration of 2–25 µg/mL in either distilled water or 10 m*M* Tris-HCl, pH 8, 1 m*M* EDTA.

2. Denature the DNA sample by heating to 95–100°C for 2–5 min in a boiling water bath, then "snap-cool" on ice.
3. Add the following to a microcentrifuge tube: DNA solution (25 ng), 1–10 µL; labeling buffer, 10 µL (solution 1); primer, 5 µL (solution 2). Water as appropriate for a final reaction vol of 50 µL; [α-^{32}P]dCTP (3000 Ci/mmol), 5 µL (50 µCi); and enzyme, 2 µL (solution 3).
4. Mix gently by pipeting up and down and cap the tube. Spin for a few seconds in a microcentrifuge to collect the contents at the bottom of the tube.
5. Incubate the reaction mix at either 37°C (for 30 min to 3 h) or at room temperature (for 3 h to overnight, *see* Note 7).
6. Stop the reaction by adding EDTA to 20 m*M*. The labeled probe can now be denatured and used directly in the hybridization or stored at –20°C. (*see* Notes 8–10).

3.7. Labeling of DNA Fragments in Low-Melting-Point Agarose

In many instances, it is preferable to use a particular segment of a DNA clone as the probe, rather than the intact clone, e.g., use of an insert fragment, free of vector sequences, can give rise to a more specific hybridization result. The following protocol can be used to label DNA directly after fractionation in low-melting-point agarose.

1. Electrophorese the restriction enzyme-digested DNA in a suitable low-melting-point agarose gel containing 0.5 µg/mL ethidium bromide. Estimate the DNA content of the desired band by reference to a set of standards on another track. Ensure that at least 250 ng of DNA is contained in the band, so that 25 ng of DNA can be used in the labeling protocol *(above)* without the need to concentrate the DNA.
2. Excise the desired band cleanly, with the minimum amount of excess agarose, and transfer to a preweighed microcentrifuge tube.
3. Add distilled water at a ratio of 3 mL/g of gel and place in a boiling water bath for 7 min to melt the gel and denature the DNA. (If the DNA is not used immediately, divide into 25-ng aliquots and store at –20°C. Reboil for only 1 min before use in the labeling reaction).
4. Transfer the tube to a 37°C water bath for at least 10 min.
5. Add the vol of DNA/agarose solution that contains 25 ng of DNA to the standard labeling reaction *(above)*. This vol should not exceed 25 µL in a 50-µL labeling reaction (*see* Note 11).

3.8. Hybridization

1. Prewarm the hybridization buffer to 65°C (*see* Note 13).
2. Immerse the blot in the hybridization buffer and prehybridize with agitation at 65°C for at least 15 min.
3. Denature the probe at 95–100°C for 2–5 min and snap-cool on ice.
4. Add the denatured probe to the hybridization buffer and mix to achieve a uniform distribution of the probe over the blot. Probe concentrations of 1–10 ng/mL are suitable for most applications (*see* Note 12).
5. Hybridize with agitation at 65°C for 2 h (if rapid hybridization buffer is used) or 16 h (for conventional buffers).
6. Wash the filter as follows:
   a. Twice in 50–100 mL of Wash 1 for 10 min at room temperature.
   b. Once in 50–100 mL of Wash 2 for 15 min at 65°C.
   c. Twice in 50–100 mL of Wash 3 for 15 min at 65°C (*see* Note 14).
7. Wrap the washed filter in Saran Wrap™ and autoradiograph (*see* Note 15).

3.9. Reprobing of Southern Blots

Following the initial hybridization, it is often desirable to remove the original probe from the blot and to "reprobe" the blot with further probes. This is especially true in cases in which the sample DNA is available in limited quantity or in applications such as population screening or fingerprinting (Chapter 22), in which a large number of hybridizations can increase the information obtained from each blot. A simple reprobing protocol is given here:

1. Boil a solution of 0.1% (w/v) SDS (*see* Note 16).
2. To remove a bound probe, pour the SDS solution onto the membrane and allow to cool to room temperature.
3. Autoradiograph to check that the probe has been removed.
4. The filter can now be prehybridized and hybridized with a new probe.

4. Notes

1. Ethidium bromide is a mutagen. Care should be taken to avoid skin contact with this reagent.
2. Lambda DNA *Hind*III fragments end-labeled with a ^{32}P- or ^{35}S-labeled nucleotide are used to provide a radioactive mol wt marker. DNA mol wt markers are available commercially.
3. Resolution of large fragments can be enhanced by performing a prolonged gel run at low voltages (e.g., for a 20 × 20-cm gel, electrophorese at 45 V overnight or 30 V for 24–48 h).

4. The xylene cyanol loading dye changes color to yellow/green and the bromophenol blue becomes yellow during depurination of the gel.
5. Avoid trapping air bubbles between the layers of the blot. If bubbles appear, they should be squeezed out using a glass rod or pipet.
6. Use the following protocol to calibrate the transilluminator:
   a. Produce six identical strips of a blot of control DNA (e.g., restricted lambda or genomic DNA). If lambda DNA is used, load 50 pg per track.
   b. Expose each blot DNA-side down on the transilluminator for different lengths of time, ranging from 30 s to 10 min.
   c. Hybridize all the blots in the same container with the same probe.
   d. Following autoradiography, the optimum UV exposure will be indicated by selecting the filter showing the strongest signal.
7. For labeling highly purified DNA (e.g., prepared by cesium chloride centrifugation), incubation times of 30 min at 37°C can be used. For lower purity DNA (e.g., DNA in agarose or DNA prepared by "minilysate" methods *[11]*), longer incubation times (3 h to overnight) are required. If incubations are carried out for longer than 3 h, they should be performed at room temperature.
8. If desired, the success of the labeling reaction can be monitored by the DEAE-paper or trichloroacetic acid precipitation methods (*see* ref. *12* for further details).
9. To obtain optimal signal-to-noise in filter hybridizations, probe purification is recommended, particularly when a labeling yields an incorporation of less than 50% of labeled nucleotide. The method of choice for probe purification is the use of "spun columns" of Sephadex-G50 (ref. *12*). Sephadex is available from Pharmacia.
10. High specific activity ^{32}P-labeled probes should be stored for no longer than 3 d.
11. The labeling reaction may appear to gel during incubation, but polymerization will still proceed if this happens.
12. A preformulated rapid hybridization buffer is available from Amersham. Hybridization times can also be reduced by the inclusion of vol excluders, such as 10% dextran sulfate in the conventional hybridization buffer. These modifications are also recommended if low probe concentrations (1–2 ng/mL) are being used.
13. A hybridization temperature of 65°C is suitable for probing DNA of an average (G + C) content (40%). The optimal hybridization and washing temperatures for probes of unusual (G + C) content will have to be determined empirically *(4)*.

14. Use a hand-held β-monitor to estimate the amount of ^{32}P bound to the filter after each washing step.
15. For ^{32}P-labeled probes, autoradiograph at –70°C using two intensifying screens and preflashed film for maximum sensitivity. For ^{35}S-labeled probes, autoradiograph dried filters at room temperature, without Saran Wrap™.
16. When using Hybond-N+, use 0.5% (w/v) SDS for probe removal.
17. If a filter is to be reprobed, do not allow it to dry completely before removing the first probe. It is extremely difficult to remove probes from dried filters.

References

1 Southern, E M (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517.

2. Mathew, C. G. P. (1984) Detection of specific sequences—the Southern Transfer, in *Methods in Molecular Biology*, vol. 2 (Walker, J. M., ed) Humana, Clifton, NJ, pp 55–66

3. Olszewska, E and Jones, K. (1988) Vacuum blotting enhances nucleic acid transfer. *Trends Genet.* **4**, 92–94

4 Meinkoth, J and Wahl, G (1984) Hybridization of nucleic acids immobilized on solid supports. *Anal. Biochem.* **138**, 267–284

5 Reed, K C and Mann, D A (1985) Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* **13**, 7207–7221.

6 Bertera, A. L , Cunningham, M. W., Evans, M. R., and Harris, D. W (1990) Filter hybridization and radiolabelling of nucleic acids, in *Advances in Gene Technology*, vol 1 (Greenaway, P. J , ed.) JAI, London, pp. 99–133.

7 Feinberg, A. P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem.* **132**, 6–13

8 Feinberg, A P. and Vogelstein, B. (1984) Addendum: A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem.* **137**, 266,267

9 Cox, K. H., DeLeon, D V., Angerer, L. M , and Angerer, R. C (1984) Detection of mRNAs in sea urchin embryos by *in situ* hybridization using asymmetric RNA probes. *Devel. Biol.* **101**, 485–502

10. Anderson, M. L. N. and Young, B. D. (1985) Quantitative filter hybridization, in *Nucleic Acid Hybridization: A Practical Approach* (Hames, B. D. and Higgins, S. J , eds.) IRL, Oxford and Washington DC, pp. 73–111.

11. Maniatis, T , Fritsch, E. F , and Sambrook, J (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

12. Rapid Hybridization System-Multiprime, protocol booklet (1988) Amersham International plc

CHAPTER 16

The Detection of Specific DNA Sequences by Enhanced Chemiluminescence

Timothy C. Richardson and Ian Durrant

1. Introduction

Blotting transfer techniques are well-established procedures for the immobilization of DNA onto solid matrices, typically nitrocellulose or nylon membranes. The Southern blotting technique (ref. *1* and Chapter 15) has found many research and, more recently, medical applications. For example, in the diagnosis of genetic diseases, such as thalassemia *(2)* and muscular dystrophy *(3)*, restriction fragment length polymorphisms (RFLPs) detected on blots are used as a basis for identifying genetic mutations (Chapter 30). In addition, the method of choice for identifying host bacteria containing recombinant DNA sequences continues to be colony or plaque screening using membrane discs.

Much of the work to date has identified specific DNA sequences on the blots by first denaturing the DNA (rendering it single stranded), and then incubating the membranes in a hybridization buffer under conditions that favor the annealing of immobilized target sequence with complementary "probe" DNA that carries a radioactive label. The radioactivity incorporated into the probe enables the presence of target sequence, on a membrane, to be identified by measuring the localized radioactive emission on X-ray film, i.e., autoradiography.

From. *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc , Clifton, NJ

Although such techniques are able to reliably detect very small amounts of target DNA, they are quite complex. The probe labeling procedure, for introducing a radioactive label, may be a lengthy procedure, and yields a reagent that, by definition, must decay and therefore has only a short working life. In addition, the small quantities of radioactivity, though not hazardous if handled correctly, do require special containment facilities, and a level of staff training, which precludes their use in certain routine laboratories.

A number of nonradioactive labeling techniques have been described, such as the incorporation of a biotinylated nucleotide into probe DNA *(4)*, or the chemical modification of particular nucleotides *(5)*. Such features introduce an internal label, that can in turn be recognized by a second, enzyme-labeled, reagent. In the case of biotin, this can be an enzyme conjugated with avidin or streptavidin. For other haptens, or chemically modified moieties, an enzyme-labeled antibody may be used. The enzyme function may then be used to generate a colored reaction product on the membrane.

Even though such procedures possess advantages by being nonradioactive, the probe labeling reactions remain quite complex, the sensitivity level of detection is often inadequate for many applications, and reprobing membranes is difficult.

We describe here a method for directly labeling probe DNA with modified horseradish peroxidase enzyme *(6)*, which is quick, nonhazardous, and produces labeled probes that can be kept for many months. Furthermore, the generation of signal by the enzyme involves a chemiluminescent reaction, which not only gives high sensitivity (down to 1pg of immobilized target, i.e., single copy genes), but also gives a "hard copy" of the result in the form of a film, which is not unlike the autoradiograph obtained using the established radioactive procedures.

The method as described applies mainly to Southern blotting, but it can equally well be applied to colony and plaque screening. However, it should be emphasized that for optimum results, the Southern blotting protocol has been very carefully analyzed, and adapted slightly, in order that this technique may be used to its maximum advantage.

2. Materials

1. The nucleic acid labeling reagents (charge-modified horseradish peroxidase and glutaraldehyde for crosslinking) and the luminol-based detection solutions that enable chemiluminescent signal generation, are available in kit form (ECL gene detection system, RPN 2101, Amersham International plc). Hybridization buffer is also available in a ready-to-use form (RPN 2102, Amersham International plc).

2. Agarose gel electrophoresis equipment is required for applications involving Southern blotting of target sequence. However, the system described is equally applicable to other membrane-based hybridization procedures (e.g., colony and plaque lifts, dot blots).
3. Solutions required, and incubation times, for gel processing prior to Southern blotting are:
   a. Depurination solution: 250 m*M* HCl; 15 min (time for dyes to change color).
   b. Denaturation solution: 1.5*M* NaCl, 0.5*M* NaOH; 30 min (time for color change plus 15 min).
   c. Neutralization solution: 1.5*M* NaCl, 0.5*M* Tris-HCl, pH 7.5; 30 min (nylon membrane), plus extra 15 min in fresh solution (nitrocellulose membrane).
   d. 20× SSC: 3*M* NaCl, 0.3*M* trisodium citrate, pH 7.5.
4. Suitable membranes are Hybond-ECL nitrocellulose and Hybond-N+ positively charged nylon (Amersham International plc) (*see* Note 1).
5. Hybridizations can be carried out in suitable boxes or chambers, or alternatively, in plastic bags placed in a shaking water bath at 42°C.
6. Posthybridization stringent wash buffer (primary wash buffer): 6*M* urea (360 g/L), 0.4% (w/v) SDS, 0.5× SSC. Prepare as a stock and store at 4°C.
7. Poststringency rinse solution (secondary wash buffer): 2× SSC.
8. Signal detection is by X-ray film. Hyperfilm-ECL (Amersham International plc) is recommended (*see* Note 2). The X-ray film cassettes need not have intensifying screens attached.
9. In many laboratories, X-ray film processing is routinely carried out in an automatic processor. X-ray film can also be processed by hand using standard developing and fixing procedures.

3. Methods

3.1. Target Preparation

1. Methodologies for the techniques of restriction enzyme digestion, agarose gel electrophoresis, and Southern blotting are described in detail in Chapter 15. Optimum results will be achieved using Hybond-ECL (nitrocellulose) or Hybond-N+ (nylon) membranes for target immobilization (*see* Notes 1 and 3).
2. Target DNA preparation for other applications can be found in many laboratory manuals *(7,8)*, in protocol booklets supplied with the membrane supports, and in Volumes 2 and 4 of this series.

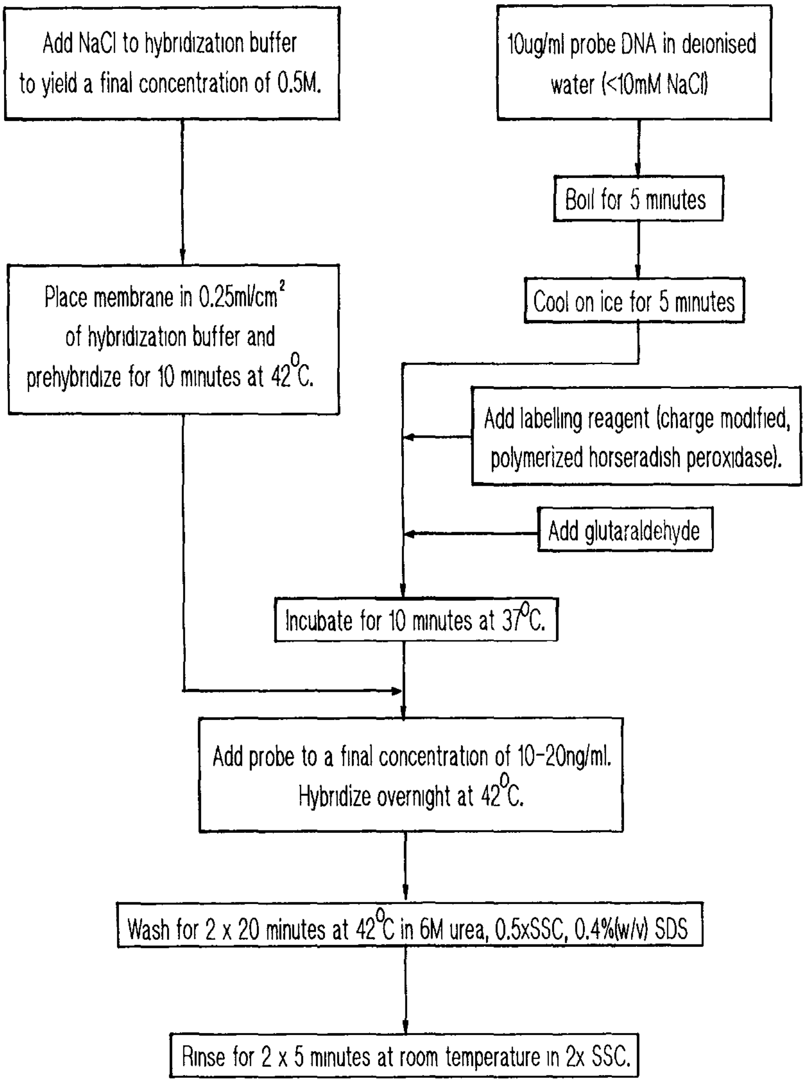


Fig. 1. An outline of the procedures required for probe labeling and hybridization using the ECL gene detection system.

3.2. Probe Labeling (see *Fig. 1*)

1. For probe preparation, dilute the DNA to a concentration of 10 ng/µL (a total of at least 200 ng is required in a vol of 20 µL; *see* Note 4). An Eppendorf tube (with cap) is a suitable container.
2. Seal the tube and boil the double-stranded DNA for 5 min in a vigorously boiling water bath (*see* Notes 5 and 6).

3. Cool the DNA on ice for 5 min.
4. Add an equal vol of DNA-labeling reagent (charge modified horseradish peroxidase *[6]*). Mix thoroughly.
5. Add a vol of glutaraldehyde (1.5% v/v), equal to that of the labeling reagent. Mix thoroughly, but briefly (1 s on a vortex mixer).
6. Centrifuge briefly (5 s) to settle the liquid at the bottom of the tube.
7. Incubate for 10 min at 37°C; *see* Note 7 for probe storage conditions.

3.3. Hybridization (see Fig. 1)

1. The hybridization buffer is available ready formulated (*see* Note 8). Sufficient buffer is required to give 0.25 mL/cm^2 of membrane. This may be halved for large blots hybridized in plastic bags.
2. Before hybridization, NaCl should be added to the hybridization buffer (*see* Note 9). When using nylon membranes, a "blocking" agent (supplied with the buffer) should also be added and fully dissolved before the buffer is used (*see* Note 10).
3. Place membranes in the hybridization buffer and incubate at 42°C for a prehybridization period of at least 15 min.
4. Add the labeled probe DNA to the hybridization buffer containing the membrane (*see* Note 11), to give a final concentration of 10–20 ng/mL (*see* Note 12). Incubate in a shaking water bath at 42°C, overnight.
5. Remove the membrane from the hybridization buffer and cover with excess primary wash buffer (*see* Note 13). Incubate in a shaking water bath for 20 min at 42°C. Repeat.
6. Place the membrane in an excess of secondary wash buffer. Incubate for 5 min at room temperature. Repeat.

3.4. Detection (see Fig. 2)

1. Using the reagents supplied in the ECL gene detection kit, mix equal vol of detection solutions 1 and 2 to give 0.125 mL/cm^2 of membrane to be developed.
2. Drain excess secondary wash buffer from the membrane filter. Lay membrane on a clean, flat surface and cover the DNA side of the membrane with freshly mixed detection reagent.
3. Incubate for precisely 1 min at room temperature.
4. Drain off excess detection reagent and wrap the filter in Saran Wrap™ (Dow Chemical Co.), ensuring that there are no creases or air pockets over the surface of the membrane. The DNA side of the membrane should be placed on the smooth side of the "parcel" facing outward.
5. Place the membrane DNA side up in an X-ray film cassette.

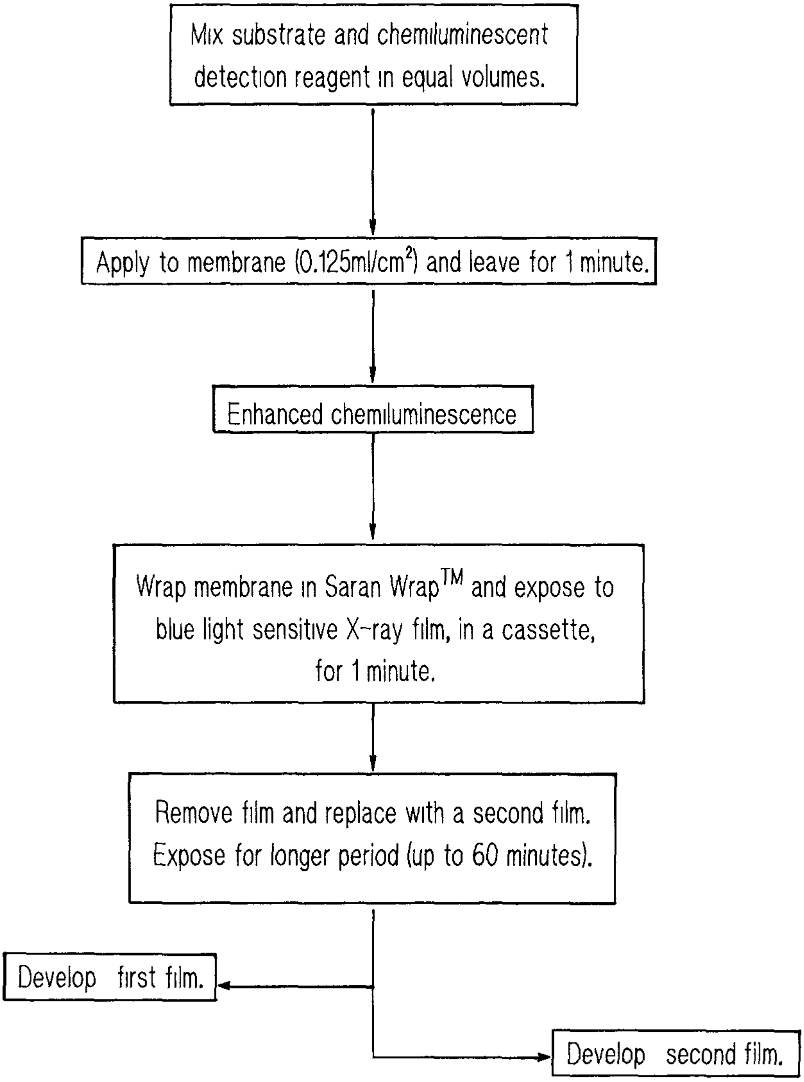


Fig. 2. The basic scheme for enhanced chemiluminescence detection of horseradish peroxidase labeled probes.

6. In a dark room, place a piece of X-ray film over the wrapped membrane and expose for exactly 1 min.
7. Remove the first film and place a second film into the cassette. Develop the first film immediately, and on the basis of the result decide how long to expose the second film (*see* Note 14). An example of the results obtainable from a Southern blot is shown in Fig. 3.

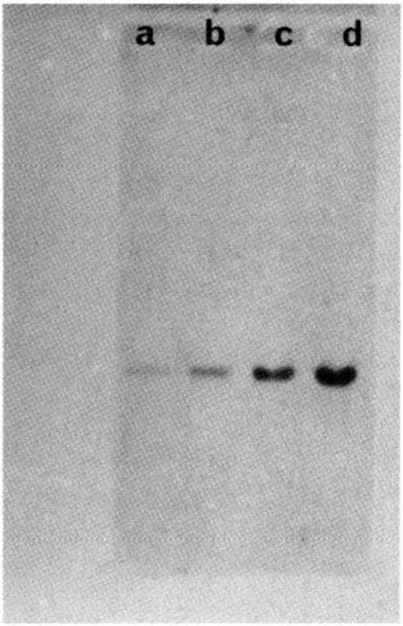


Fig. 3. Single copy gene detection in genomic DNA. EcoRI-restricted human genomic DNA, blotted onto Hybond N+, probed with a 1.5-kb fragment of the N-*ras* protooncogene sequence. Labeling and hybridization were performed as described in the text. Gel loadings of 1, 2, 5, and 10 µg in lanes a–d, respectively. Probe concentration 20 ng/mL; film exposure of 15 min.

8. Following this whole procedure, membranes may be reprobed (*see* Note 15). Nitrocellulose may be reprobed up to 5 times, and positively charged nylon at least 10 times. The limiting factors are membrane damage caused by repeated handling and gradual loss of target. Between probings, the membranes should be kept moist by storage in the Saran Wrap™ "parcel."

4. Notes

1. It is possible to obtain nitrocellulose and nylon membranes from a number of suppliers. However, the ECL gene detection system has been optimized for use on Hybond-ECL (nitrocellulose) and Hybond N+ (nylon). Membranes from other sources may not yield the maximum sensitivity.
2. Although Hyperfilm-ECL is recommended, Hyperfilm-MP and Kodak X-ray film may give acceptable results. However, blue-tinted films are not recommended for best results.

3. Target DNA preparation is of particular importance to the subsequent successful use of the ECL gene detection system. Restriction enzyme digests must be performed to completion; ideally an aliquot should be checked on an agarose minigel prior to loading the experimental gel. The electrophoretic separation must be sufficient to resolve the size of DNA bands expected, and the postelectrophoresis treatments of the gel should follow the recommended protocol wherever possible.
4. DNA to be labeled must be dissolved in a solution containing less than 10 m*M* monovalent cations. At higher concentrations, labeling efficiency will decrease owing to incomplete denaturation of double-stranded probes and inhibition of the electrostatic interactions between the labeling reagent and the nucleic acid. In order to achieve the optimum specific signal over and above background (i.e., maximum ratio of signal:noise), the DNA should ideally be an insert sequence excized and purified from the host vector. Most of the standard insert purification procedures will produce probe DNA that is suitable for use with the ECL gene detection system. Additionally, probes may be labeled directly from a 0.7% (w/v) low-gelling temperature agarose gel slice.
5. Single-stranded DNA can be labeled, but a boiling step is not required before the addition of the labeling reagent. RNA probes can also be constructed; neither do these need denaturation by boiling. However, for RNA probes, the vector DNA sequences must be digested with DNase I (RNase-free) prior to labeling.
6. Heating blocks or water baths at 95°C are not suitable for the complete denaturation of probe DNA necessary to achieve maximum labeling.
7. Labeled probes can be stored on ice for 15–30 min prior to use. For longer storage, up to 6 mo, an equal vol of glycerol should be added, and the solution stored at –20°C.
8. Some of the components of the hybridization buffer may come out of solution during storage; in particular, there is a high concentration of urea, which is included as a helix destabilizing agent. The crystals should be redissolved by warming the buffer to 65°C, with shaking, over a 15–30 min period; this procedure is not detrimental to the performance of the buffer, providing that the temperature of 65°C is not exceeded. NaCl can be added to the buffer at this stage if desired (*see* Note 9), and the buffer can be aliquotted (e.g., into 25 mL Universal containers) and stored frozen at –20°C. When thawed, the contents of the aliquots readily redissolve.
9. The stringency of the hybridization cannot be controlled by increasing the temperature above 42°C owing to the thermal instability of the

horseradish peroxidase during an overnight hybridization. Stringency may be controlled by NaCl concentration and the buffer is produced without NaCl to enable this parameter to be varied. If a suitable NaCl concentration has not been determined empirically, then 0.5*M* will be suitable for most applications, offering maximum hybridization for homologous sequences. With mismatched probes, it may be necessary to alter the salt concentration in the range 0.5–1.0*M* or to reduce the hybridization temperature.

10. The use of positively charged nylon membrane, which has a greater inherent protein binding capacity, requires the use of a blocking agent in the hybridization buffer. This is added to 5% (w/v) final concentration. However, it is not readily soluble and the buffer should be heated to 65°C for up to 1 h, with vigorous agitation, to ensure complete dissolution. Aliquots of hybridization buffer with blocker present can be stored at –20°C (*see* Note 8) for 3 mo.
11. Adding the nucleic acid probe directly onto the membrane should be avoided, otherwise there may be a local area of high nonspecific binding. Some of the hybridization buffer may be removed from the hybridization vessel for mixing with the labeled probe, and the mixture then returned to the bulk of the buffer. The hybridization buffer does not need to be changed between prehybridization and hybridization.
12. In general, a probe concentration of 10 ng/mL is optimal for nylon membranes and 20 ng/mL for nitrocellulose membranes. (The difference is caused by the increased target retention associated with nylon membrane.) However, for applications that have a high target level (for example, colony and plaque screening), it may be possible to halve the above probe concentrations.
13. As for hybridization, the stringency of the washing step may be controlled by NaCl concentration but not by increasing the temperature above 42°C. In addition, stringency may be altered by changes in the urea concentration of the primary wash buffer. Basically, the SSC concentration can be altered in the range 0.1–0.5× to increase stringency and from 0.5–2× to decrease stringency. Decreasing the urea concentration in the range 1–6*M* or decreasing the temperature will also decrease the stringency. Such changes would have to be determined experimentally for individual probes as necessary.
14. The exposure time for the second film may be anything from 10–60 min. This depends mainly on the concentration of target DNA molecules on the membrane, and also on the background binding as influenced by the purity of the probe and the stringency of the hybridization and

primary wash. In general, nylon-based systems offer greater sensitivity, despite higher backgrounds, relative to nitrocellulose membranes, because of increased target retention on the membrane. Consequently, shorter film exposures may be used to obtain a similar level of sensitivity to that seen on a comparable nitrocellulose membrane.

15. Reprobing membranes is extremely simple with the ECL gene detection system. Unlike other systems (both radioactive and nonradioactive), there is no requirement to remove hybridized probe from the membrane in order to carry out a reprobing. The enhanced chemiluminescent reaction leads to the inactivation of the horseradish peroxidase label (after 4–5 h). Experiments suggest that ensuing hybridizations cause some strand displacement of the previous probe. In addition, subsequent probes (even if they are the same as the previous one) can hybridize to target sequence that was not covered during the first hybridization. For nylon membranes, each subsequent hybridization requires the use of hybridization buffer containing the blocking agent.

References

1. Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol. Biol.* **98,** 503–517
2. Thein, S. L. and Weatherall, D. J. (1987) Approach to the diagnosis of beta-thalassemia by DNA analysis *Acta Haematol. (Basel)* **78,** 159–167
3. Goodship, J., Malcolm, S , Robertson, M. E., and Pembrey, M E. (1988) Service experience using DNA analysis for genetic prediction in Duchenne muscular dystrophy. *J Med. Genet.* **25,** 14–19.
4. Langer, P. R., Waldrop, A. A., and Ward, D. C. (1981) Enzymatic synthesis of biotin-labeled polynucleotides: Novel nucleic acid affinity probes. *Proc. Natl. Acad. Sci. USA* **78,** 6633–6637.
5. Verdlov, E. D., Monastyrskaya, G. S., Guskova, L. I , Levitan, T. L., Sheichenko, V. I., and Budowsky, E. I. (1974) Modification of cytidine residues with a bisulfite-O-methylhydroxylamine mixture. *Biochim. Biophys Acta* **340,** 153–165.
6. Renz, M. and Kurz, C. (1984) A colorimetric method for DNA hybridization *Nucleic Acids Res.* **12,** 3435–3444.
7. Maniaitis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
8. Berger, S. L. and Kimmel, A. R., eds. (1987) *Guide to Molecular Cloning Techniques: Methods in Enzymology,* vol. 152. Academic, New York

CHAPTER 17

Pulsed-Field Gel Electrophoresis

Johan T. den Dunnen
and Gert-Jan B. van Ommen

1. Introduction

Conventional agarose gel electrophoresis is capable of separating DNA fragments with sizes of up to 20–30 kbp. In 1984, Schwarz and Cantor *(1)* developed an electrophoretic technique capable of resolving DNA molecules in excess of 2,000,000 base pairs (2.0 Mbp). They called the technique pulsed-field gradient gel electrophoresis. Its basic principle is a continuous reorientation of the DNA molecules, caused by a recurrent change in electric field direction. This results in a migration velocity in the net field direction, depending primarily on the size of the DNA molecules.

Later, similar techniques were described, all with modifications on the principle of DNA reorientation. Variations were tested of the electrode configuration, the polarity, or the position of the gel in the box. For instance, Olson and collaborators first described orthogonal-field-alternation gel electrophoresis *(2)* and later field-inversion gel electrophoresis *(3)*. In the latter system, the field polarity is simply reversed in alternating switching intervals with a 3/1 ratio of forward to reverse fields. The term PFGE is nowadays used as an acronym for pulsed-field gel electrophoresis to indicate any technique that resolves DNA by continuous reorientation.

Field-inversion gel electrophoresis (FIGE, ref. *3*) and contour-clamped homogeneous electric field electrophoresis *(4)* (CHEF) are the most commonly used PFGE systems. They will be described in detail in this chapter. Other frequently used systems include the "Waltzer" *(5)*, wherein the gel lies

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

on a turntable, and transverse alternating-field electrophoresis (TAFE, ref. *6*), which contains a gel vertically placed, perpendicular to the electric field.

Theoretical considerations still do not fully explain all electrophoretic phenomena that are observed with the individual systems. Several studies have been performed to improve our insight into the way in which the DNA resolution is achieved by the continuous reorientation of the DNA molecules *(7,8)*. Other studies have led to the construction of computer models that generate theoretical mobility curves. These models can be used to derive the parameter settings to be used for the optimal separation in a desired size range *(9,10)*.

We have previously described the methodology of analyzing human DNA by PFGE *(11)*. The two systems that we currently use to study the Duchenne Muscular Dystrophy (DMD) gene *(12,13)* are the FIGE and CHEF systems. This chapter describes their use in combination with a commercially available power supply that provides a programmable, recurrent inversion of output polarity.

The availability of DNA of very high mol wt (larger than 5 Mbp) is essential for the successful utilization of PFGE. Standard DNA-isolation protocols cannot be used. They lead to mechanical shearing of the DNA to molecules smaller than 200 kbp. The simplest way to circumvent this problem is an encapsulation of the cells in agarose prior to cell lysis *(11,14)*.

Furthermore, PFGE requires the use of specific, infrequently cutting restriction endonucleases (rare-cutters), a modification of the protocols to digest the agarose-embedded DNA, altered techniques to load the DNA samples onto a gel, the preparation and use of DNA marker molecules in the size range of over 50 kbp, and a modification of the techniques to blot and hybridize the DNA. This chapter supplies the essential protocols, and the application of PFGE is described in Chapter 26.

2. Materials

1. Perspex mold: a perspex block-former containing rectangular holes of $10 \times 6 \times 1.5$ mm (Fig. 1).
2. Nylon membrane; Bio-Trace^RP (Gelman Sciences Inc.). Other membranes, such as Hybond-N-Plus™ (Amersham) and GeneScreen-Plus™ (NEN), can also be used.
3. Electrophoresis box: FIGE: Electrophoresis is done in a standard horizontal submarine gel box (Fig. 2) that allows circulation of the buffer. The gel rests on a table and is secured at each end with two pegs.

   CHEF: Electrophoresis is done in a rectangular gel box (Fig. 2). The electrodes are fixed, in a hexagonal configuration to the lid of the gel box (Fig. 2, cf Chu et al. *[4]*). The gel rests on a table and is secured at

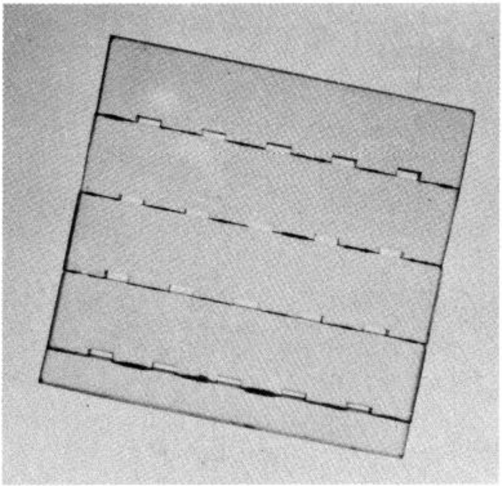


Fig. 1. Perspex mold to prepare agarose plugs. The mold was constructed from 10-mm thick perspex strips of 10 × 2 cm, in which slits 6 mm wide and 1.5 mm deep were made on one side. The strips were then glued together to form the slots.

each corner with two pegs. A practical design has recently been published *(15)*. Commercial systems are available from LKB and Bio-Rad.

Both systems: Electrophoresis buffer is cooled to 18°C and circulated through the container. To assure even cooling during electrophoresis, the gel is covered with a perspex plate that has the same thickness as the table on which it rests. This method allows more gels to be stacked together and run simultaneously.

4. Power supply: The system we describe is based on the use of the Gene-Tic™ (Biocent, P.O. Box 280, 2160AG Lisse, The Netherlands). This power supply has the described program built in and is capable of driving either four FIGE or two CHEF gels in parallel, although each gel can be programmed independently. Other commercially available power supplies and switch devices lack one or more of the possibilities mentioned, but can be applied with adaptations.
5. Agarose; InCert-agarose (FMC) is used for the isolation of DNA which has to be digested. Low gelling temperature (LGT)-agarose (BRL or Bio-Rad) is used for marker DNA isolations. SeaKem LE-(FMC) or Sigma low-electroendosmosis (EEO)-agarose (A6013) are used for gel electrophoresis.
6. Blood lysis buffer: 155 mM NH_4Cl, 10 mM $KHCO_3$, 1 mM ethylene diamine tetra acetic acid (EDTA).

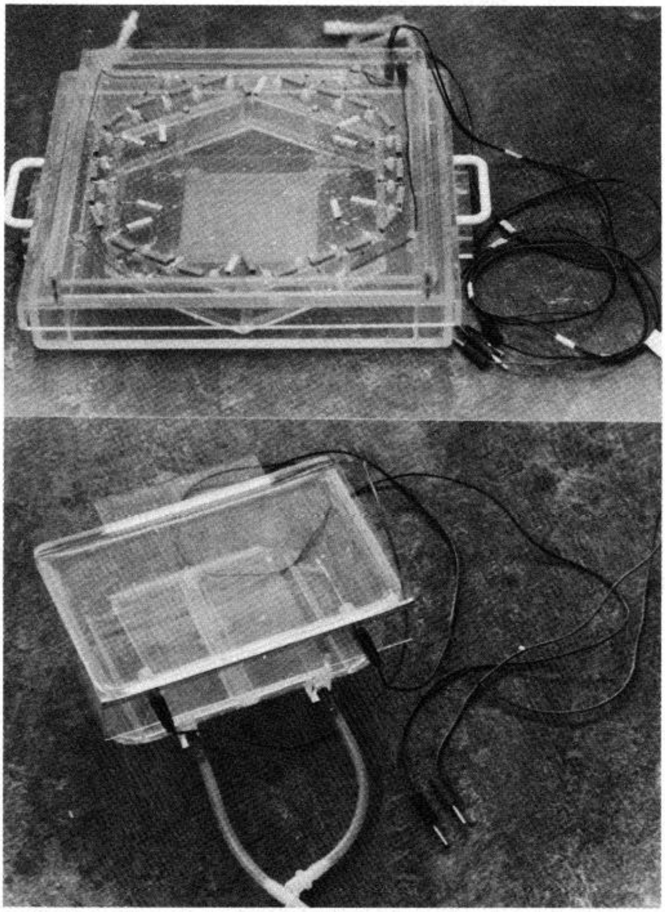


Fig. 2. Photograph of the CHEF (top) and FIGE boxes (bottom) used. A detailed description is given in the text (Section 2).

7. Competitor DNA: 500 ng/mL placenta DNA sonicated to 100–1000 bp.
8. ES: 0.5*M* EDTA, 1% Na-*N*-lauroylsarcosinate (Sigma), pH 9.5.
9. Electrophoresis buffer: 45 m*M* Tris-HCl, 45 m*M* boric acid, 0.5 m*M* EDTA, pH 8.3. Store as a 20× concentrated stock solution.
10. Equilibration buffer: Enzyme-specific restriction endonuclease incubation buffer, made as recommended by the manufacturer, containing 2 m*M* spermidine, but lacking bovine serum albumin (BSA). Store at –20°C as a 10× concentrated stock solution.

11. Ethidium bromide solution: 0.5 µg/mL ethidium bromide in H_2O. Store as a 10 mg/mL stock solution.
12. HYB solution: 0.125*M* Na_2HPO_4 (pH 7.2 with H_3PO_4), 0.25*M* NaCl, 1.0 m*M* EDTA, 7% sodium dodecyl sulfate (SDS) (BDH 44244), 10% polyethylene glycol (PEG)-6000 (BDH).
13. Neutralizing buffer: 1.5*M* NaCl, 0.5*M* Tris-HCl, pH 7.0.
14. PMSF: phenylmethylsulfonyfluoride (Sigma, P7626).
15. SE: 75 m*M* NaCl, 25 m*M* EDTA, pH 7.5.
16. SED: 75 m*M* NaCl, 25 m*M* EDTA, pH 8.0, 20 m*M* 1,4-di thio threitol (DTT).
17. SSC: 150 m*M* NaCl, 15 m*M* sodium citrate (pH 7.0). Stored as a 20× stock solution.
18. TE: 10 m*M* Tris-HCl, 1.0 m*M* EDTA, pH 7.5.
19. YPD: 1% yeast extract, 2% peptone, 2% dextrose.
20. Zymolyase: Zymolyase-20T, Seikagaku Kogyo Co. Ltd., Tokyo, Japan.
21. Nuclease free BSA. Prepare a stock solution at 2 mg/mL.

3. Methods

Definition: A plug is a 100-µL 0.5% agarose block (10 × 6 × 1.5 mm) containing DNA.

3.1. DNA Isolation in Agarose Blocks (11)

DNA is isolated from white blood cells. On average, 10 mL of blood yields enough leukocytes to prepare about 20 plugs (*see* Note 1).

1. Take 10 mL of heparinized blood and add 30 mL of blood lysis buffer. Leave for 15 min on ice to ensure complete hemolysis by isotonic ammonia treatment. Centrifuge the white cells for 15 min at 2000*g*.
2. Resuspend the pellet in 10 mL of blood lysis buffer, leave for 15 min on ice, and centrifuge for 15 min at 2000*g*.
3. Resuspend the cells thoroughly at 20×10^6 cells/mL in SE. Mix in a 1/1 ratio with 1% InCert-agarose in SE, cooled to 50°C.
4. Dispense the mixture immediately into the slots of a perspex mold (Fig. 1) covered on one side with tape. Place the mold on ice for 5–10 min.
5. Remove the tape and gently blow the solidified blocks out of the slots, using a Pasteur pipet balloon, into 5 vol of ES containing 0.5 mg/mL pronase (preincubated for 1 h). Incubate overnight at room temperature under gentle rotation (*see* Note 2).
6. Rinse the plugs once with sterile water, and wash, four times for 2 h each and once overnight, in 10–20 vol of TE, under gentle rotation (*see* Note 3).
7. Store the plugs in 0.5*M* EDTA, pH 8.0, at 4°C (*see* Note 4).

Table 1
Sizes of PFGE Marker Molecules[a]

λ	*S. cerevisiae*	*C albicans*	*S. pombe*
–	>2.3	>2.5	5.7
0.728	1.45	>2.3	4.6
0.679	1.20	2.15	3.5
0.631	0.97	1.80	–
0.582	0.94	1.63	–
0.534	0.82	1.60	–
0.485	0.79	1.20	–
0.437	0.75	1.08	–
0.388	0.68	0.97	–
0.340	0.60[b]	–	–
0.291	0.44	–	–
0.243	0.36	–	–
0.194	0.28	–	–
0.146	0.23	–	–
0.097	–	–	–
0.049	–	–	–

[a]Bacteriophage λ (CI857Sam7) has a genome size of 48.5 kbp. Yeast strains used are *Saccharomyces cerevisiae* AB1380 *(17)*, *Candida albicans* CBS562 *(18)* and *Schizosaccharomyces pombe* CBS356 *(18)*. Yeast chromosome sizes are callibrated with the λ ladder. Sizes are given in megabase pairs (Mbp), starting with the largest molecule.

[b]Doublet band

3.2. Preparation of Bacteriophage λ Marker Plugs

Use DNA of λ CI857Sam7, which has a genome of 48.5 kbp (Table 1).

1. Dilute the bacteriophage λ DNA to 5–10 μg DNA/mL in SE. Mix 1/1 with 1% LGT-agarose in SE cooled to 50°C. Dispense into the slots of a perspex mold and allow to solidify on ice for 5–10 min.
2. Proceed as in Section 3.1, Step 5 (*see* Notes 4–6).

3.3. Preparation of Yeast Marker Plugs

Routinely, 20 mL of *Saccharomyces cerevisiae* culture is used to prepare 40–100 plugs. Chromosome sizes are given in Table 1 (*see* Note 7).

1. Inoculate 20 mL of YPD and grow overnight at 37°C, under vigorous shaking, to late log phase.
2. Collect the cells by centrifugation for 10 min at 1500*g*. Wash the cells in 50 m*M* EDTA, pH 8.0, and centrifuge again.

3. Resuspend the cells in SED. Add Zymolyase-20T to total 30 µg/mL, mix 1/1 with 1% LGT-agarose in SE cooled to 50°C, and dispense immediately into the slots of a perspex mold covered on one side with tape. Place on ice for 5–10 min (*see* Note 8).
4. Remove the tape and gently blow the solidified blocks out of the slots, using a Pasteur pipet balloon, into 2 vol of SED with 30 µg/mL Zymolyase-20T. Incubate for 1–2 h at 37°C under gentle rotation.
5. Rinse the plugs once with SE. Transfer the plugs to 2 vol of SE containing 1.0 mg/mL pronase (preincubated for 1 h), and incubate overnight at room temperature, under gentle rotation.
6. Proceed as in Section 3.1.6.

3.4. Restriction Endonuclease Digestion of DNA in Agarose Blocks

Newly prepared plugs of mammalian DNA should be checked for residual nuclease contamination by a control incubation without the addition of enzyme (*see below*). After incubation, the DNA is analyzed on a PFGE gel. DNA degradation should be negligible in the size range under study, i.e., up to at least 2 Mbp (*see* Note 9). Usually, half-plugs are digested and loaded into each lane (equaling 5–7.5 × 10^5 cells, or 3–5 µg of DNA).

1. Rinse the plugs once with sterile water and wash three times for 30 min each in 10–20 vol of TE under gentle rotation.
2. Place a half-plug in 1.0 mL of equilibration buffer and incubate for 2 h at room temperature, or overnight at 4°C.
3. Carefully remove all the equilibration buffer. Add 50 µL of fresh equilibration buffer containing 0.2 mg/mL BSA. Digest for 6 h (or overnight) at the specified incubation temperature, using 15–25 U of enzyme. Add the enzyme in two equal portions at the beginning, and after 3 h of digestion (*see* Notes 10–12).
4. After digestion, place on ice for 15–30 min. Remove all buffer.
5. The plugs may be either used directly or stored in 50 m*M* EDTA, pH 8.0, at 4°C (*see* Note 13) for later use.

3.5. Pulsed-Field Gel Electrophoresis: The Gene-Tic

Figure 3 shows an example of the possibilities of a gel separation for both FIGE and CHEF system. In both systems, the size range over which the DNA is separated is defined by the parameter settings of a file that drives the electrophoresis. We define the interval between subsequent inversions of the electrode polarity as the "switch time." For FIGE, each run is divided into

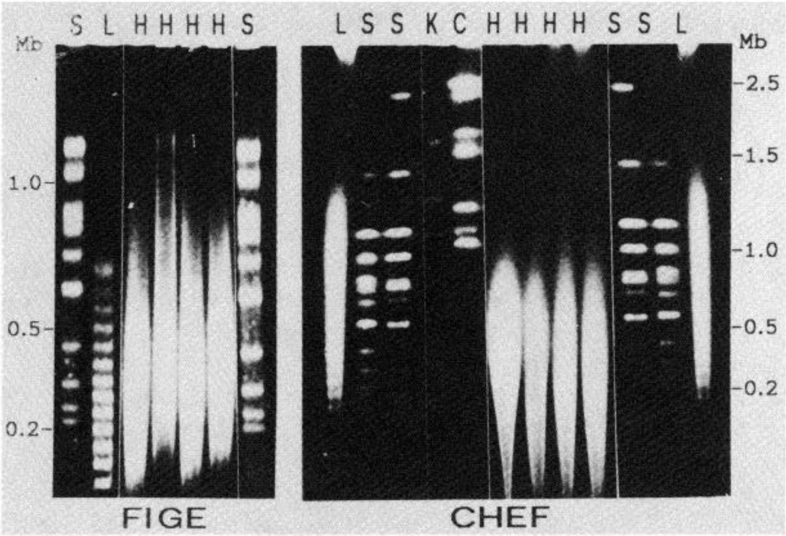


Fig. 3. Photograph of ethidium bromide stained FIGE (left) and CHEF gel (right). The FIGE gel represents a standard electrophoretic separation; 7.5 V/cm for 18 h at 18°C with a 40% exponential switch time increase from 1–60 s in four identical cycles and a 2% pause interval *(see text)*. The CHEF gel was run to separate DNA molecules up to 2.5 Mbp: Electrophoresis was at 3.2 V/cm for 68 h at 18°C with a linear switch time increase from 1 to 500 s and a 2% pause interval. Sizes are indicated in Mbp. DNAs used: L = bacteriophage λ, S = *S. cerevisiae*, C = *C. albicans*, K = *K. lactis*, and H = SfiI-digested human DNA.

cycles of 4–6 h (Fig. 4). The shortest switch time, at the start of each cycle, defines the lower limit of separation; the longest switch time, at the end of each cycle, defines the upper limit of DNA molecules that are resolved. We have introduced a short pause each time the electric field is reversed. This allows a relaxation of the DNA molecules and was found to result in an improved resolution above 400 kbp. The time for electrophoresis in the backward direction is set to 1/3 of that in the forward direction. An exponential mode of switch-time increase is available, in which the user is requested to set the percentage of the total cycle duration at which 50% of the switch-time increase is reached. A figure of 40% provides a time-ramp curve that initially increases more rapidly (Fig. 4). This setting was found to improve markedly the separation of the larger DNA molecules.

CHEF is usually performed at a constant switch time (Fig. 4), consequently in one linear cycle. The electric field is not reversed between two electrodes, but alternates in polarity between two sets of electrodes to give a 120° reorientation of field angle.

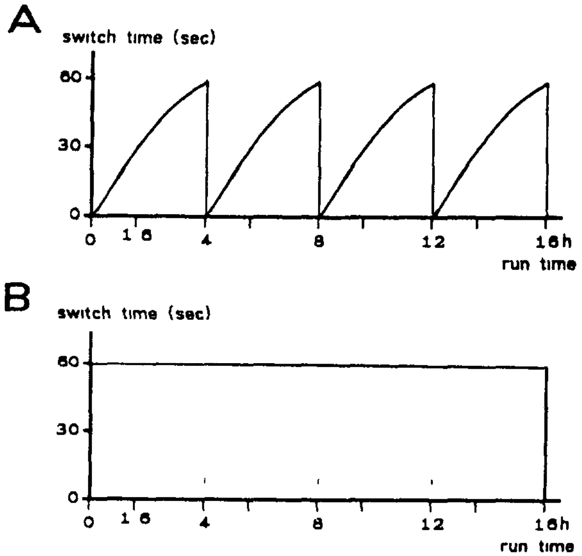


Fig. 4. Graphical illustration of the length of the switch time during (A) a standard FIGE-run (Fig. 3) or (B) a standard CHEF-run (*see* Section 3.5).

3.6. Running PFGE Gels

1. Prepare a 1% agarose gel by adding agarose powder to electrophoresis buffer. Boil until the solution is clear. Cool the agarose solution to 60°C and pour it into a gel mold, insert a well-former, and allow the gel to set for 45–60 min.
2. Carefully remove the well-former and load the gel on the laboratory bench. Load bacteriophage λ and yeast-marker plugs by inserting them directly into the slots.
3. Plugs containing digested DNA are melted for 10 min at 65°C and then carefully layered from the side of a well (to avoid air bubbles), using a 200-µL-capacity micropipet tip from which the last 5–8 mm have been cut off.
4. Fill the electrophoresis box with electrophoresis buffer, carefully submerge the loaded gel, and cover with a perspex plate. Turn on the cooling and leave the gel for 30 min.
5. Electrophorese using the Gene-Tic power supply with parameter settings for a DNA separation in the desired size range (*see* Note 14). For a standard DNA separation from 30–1000 kbp, the following parameters are used:

a. FIGE: 7.5 V/cm for 15 h at 18°C with four identical cycles, each with a switch interval increasing from 1 s at the beginning to 60 s at the end. The time ramp increases exponentially in such a way that 50% of the switch-interval increase is reached at 40% of each cycle duration (Fig. 4). The reverse switch interval measures 33% of the preceding forward one. A pause interval of 2% of the forward switch time is included.
b. CHEF: 7.5 V/cm for 18 h at 18°C, usually with a constant switch interval of 60 s and a pause interval of 2% (Fig. 4).

6. For gels to be blotted, *see* Section 3.7. Stain analytical gels for 30–60 min in an ethidium bromide solution. Photograph gels on a UV-transilluminator, either directly or after improvement of the contrast by washing for 1–2 h in several changes of H_2O.

3.7. Blotting of PFGE Gels

1. After electrophoresis, stain the gel for 30–60 min in ethidium bromide solution. Photograph on a UV transilluminator, either immediately or after improvement of the contrast by extensive H_2O-washing.
2. Reduce DNA size by UV irradiation, either for 60–90 s with 254 nm light or for 5–10 min with 302 nm light (*see* Note 15).
3. Wash the gel twice for 15 min each time in 0.4*M* NaOH. Rinse with water, wash once for 20 min in neutralizing buffer and subsequently once for 20 min in 10× SSC. Blot the gel upside down (*see* Note 16), at least overnight in 10× SSC onto a nylon membrane.

3.8. Hybridization of PFGE Blots

1. Label 10 ng of probe DNA with ^{32}P-α-dCTP using a random oligonucleotide labeling reaction (e.g., Multiprime kit, Amersham). Remove unincorporated nucleotides by purification over a Sephadex G50 column in a Pasteur pipet.
2. Prehybridize the blots in HYB solution for at least 10 min at 65°C.
3. Add labeled probe to the prehybridization, mix thoroughly, and hybridize overnight at 65°C.
4. Wash the blots at 65°C in 2.0× SSC/0.1%SDS (2×, 15 min each), 1.0× SSC/0.1%SDS (2×, 15 min each), and finally to 0.3× SSC/0.1%SDS (once for 15min).
5. Autoradiograph with Kodak™ X-Omat R film overnight (or longer if required) at –70°C, using intensifying screens (Dupont).

3.9. Competitive DNA Hybridization of PFGE Blots (16)

Note: This step is necessary if the probe contains repeated sequences.

1. Label 10 ng of probe DNA with ^{32}P-α-dCTP (Section 3.8, Step 1).
2. Prehybridize the blots in HYB solution for at least 10 min at 65°C.
3. Transfer half of the labeling reaction (ca. 200 µL) to an Eppendorf tube. Add 240 µL competitor DNA (ca. 20 × 10^3 excess), boil for 5 min, and chill on ice. Add to 1.5 mL of HYB solution (preheated to 65°C), mix thoroughly, and incubate for 90 min (N.B.: Time is crucial!) at 65°C in a water bath.
4. Add the mixture to the prehybridization, mix thoroughly, and incubate overnight at 65°C in a water bath.
5. Further handling is as described in Section 3.8, steps 4 and 5.

3.10. Rehybridization of PFGE Blots

1. Boil 200 mL of 0.1× SSC and pour into a tray. Immediately add the used blots, cover the tray, and leave for 3 min with gentle shaking.
2. Take the blots out and put them into a new tray containing 2× SSC/0.2*M* Tris-HCl (pH 7.5) and leave for 5 min.
3. Air-dry the blots. The blots are now ready for new hybridizations.

4. Notes

1. Heparinized blood can be stored at –70°C before DNA isolation is done. DNA isolation from other sources, such as tissue-culture cells, sperm cells (add 10 m*M* DTT in Section 3.1, steps 2 and 3), and fresh or frozen tissues (after homogenization to a single-cell suspension) is also possible using the same protocol.
2. Pronase routinely gives satisfactory results. Proteinase K (0.5 mg/mL, incubation overnight at 50°C) can be used instead, but is more expensive.
3. Addition of 40 µg/mL PMSF in the first two TE-washing steps can be used to reduce protease activity.
4. Storage in ES or TE is also possible. Storage in TE is dangerous (there is a high risk of DNA degradation after minor nuclease contaminations, e.g., from poorly digested cells), but allows digestions to be started without extensive washings (*see* Section 3.4, Step 2)
5. Preparation of λ marker plugs from isolated phage particles may be preferred, because commercial DNA preparations give variable results.
6. Annealing of the λ sticky ends depends mainly on the DNA concentration in the plugs and the temperature during preparation. When the

ladders do not reach the desired size range, they can be enlarged by incubation in $MgCl_2$; equilibrate the plugs to 10 m*M* $MgCl_2$ and incubate for 15 min at 42°C. Wash extensively in TE. Store in 0.5*M* EDTA (pH 8.0).

7. The chromosome sizes obtained differ between yeast strains. The yeast strains used are described in Table 1.
8. A smear throughout the lanes after electrophoresis indicates poor cell lysis, DNA degradation or RNA contamination. Cell lysis can also be done with Novozym (SP234, NOVO Industri AS, Copenhagen, Denmark) or lyticase (Sigma, L5263). A dominant smear of RNA, obscuring chromosome bands, can be removed by a RNase treatment.
9. Degraded or broken DNA can be removed from a plug by a short "pre-electrophoresis" before further handling of the sample. Remaining DNA-degrading activities can be removed by a second pronase treatment.
10. Frequently used rare-cutter enzymes are SfiI, SalI, SacII, BssHII, MluI, NotI, NarI, NruI, and NaeI.
11. For double digestions: repeat Steps 1–3 in Section 3.4 for the second enzyme. Before Step 1 of the second digestion, a proteinase K treatment (0.5 mg/mL) can be inserted (this is not essential).
12. For digestions with normal restriction endonucleases (EcoRI, HindIII, and so on) modify Step 3 in Section 3.4 to the following: Carefully remove all equilibration buffer and incubate for 10 min at 65°C to melt the plug. Incubate for 15 min at 37°C, add BSA (to 0.1 mg/mL), and add restriction enzyme. Incubate at the desired temperature. Layer the molten plug directly onto the gel.
13. When fragments larger than 1.5 Mbp are to be detected, plugs are not melted, but layered directly onto the gel, to prevent shear.
14. Under the given conditions, a rule of thumb is that an increase by 1 s of the forward switch time at the end of a cycle results in an upward shift of the zone of unseparated DNA by 20 kbp.
15. It is advisable to test each UV transilluminator to define the optimal illumination time. Usage of the UV Cross-Linker (Stratagene), irradiating to a preset intensity, may be preferred. Size reduction by acid depurination (20 min of incubation in 0.25*M* HCl, rinse in water, wash for 20 min in neutralizing buffer, wash twice for 20 min each time in 10× SSC) is possible, but gives variable results.
16. Upside-down blotting prevents occasional variations in transfer efficiency, caused by "skin formation" when agarose solutions were standing too long before gels were poured.

Acknowledgments

We gratefully acknowledge the skillful assistance of W. F. A. Bingley, R. D. Runia, and L. Gerrese in the construction and modifications of the FIGE and CHEF boxes and the electronic equipment and J. M. H. Verkerk and M. Rinkels for their technical assistance with setting up the system. This work was supported in part by grants from the Dutch Prevention Fund, the Netherlands Scientific Research Organisation, the Muscular Dystrophy Group of Great Britain, and the Muscular Dystrophy Association of America.

References

1. Schwarz, D. C. and Cantor, C. R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75.
2. Carle, G. R. and Olson, M. V. (1984) Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. *Nucleic Acids Res.* **12**, 5647–5664.
3. Carle, G. R., Frank, M., and Olson, M. V. (1986) Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. *Science* **232**, 65–68.
4. Chu, G., Vollrath, D., and Davis, R. W. (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* **234**, 1582–1585.
5. Southern, E. M., Anand, R., Brown, W. R., and Fletcher, D. S. (1987) A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nucleic Acids Res.* **15**, 5925–5943.
6. Stewart, G., Furst, A., and Avdalovic, N. (1988) Transverse Alternating Field Electrophoresis (TAFE) *BioTechniques* **6**, 68–73.
7. Schwartz, D. C. and Koval, M. (1989) Conformational dynamics of individual DNA molecules during gel electrophoresis. *Nature* **338**, 520–522.
8. Smith, S. B., Aldridge, P. K., and Callis, J. B. (1989) Observation of individual DNA molecules undergoing gel electrophoresis. *Science* **243**, 203–206.
9. Lalande, M., Noolandi, J., Turmel, C., Rousseau, J., and Slater, G. W. (1987) Pulsed-field electrophoresis: application of a computer model to the separation of large DNA molecules. *Proc Natl. Acad. Sci USA* **84**, 8011–8015.
10. Heller, C. and Pohl, F. M. (1989) A systematic study of field inversion gel electrophoresis. *Nucleic Acids Res.* **17**, 5989–6003.
11. van Ommen, G J. B. and Verkerk, J. M. H. (1986) Restriction analysis of chromosomal DNA in a size range up to two million base pairs by pulsed field gradient electrophoresis, in *Human Genetic Disease, A Practical Approach* (Davis, K. E., ed), IRL, Oxford, pp. 113–133.
12. den Dunnen, J. T., Bakker, E., Klein-Breteler, E. G., Pearson, P. L., and G. J. B. van Ommen. (1987) Direct detection of more than 50% Duchenne muscular dystrophy mutations by field inversion gels. *Nature* **329**, 640–642.
13. den Dunnen, J. T., Bakker, E., van Ommen, G. J. B., and Pearson, P. L. (1989) The DMD gene analysed by field inversion gel electrophoresis. *Br. Med. Bull.* **45**, 644–658.

14. Smith, C. and Cantor, C. R. (1987) Purification, specific fragmentation and separation of large DNA molecules, in *Methods in Enzymology, Recombinant DNA*, vol. 155 (Wu, R., ed.), Academic, London, pp. 449–467.
15. Meese, E. and Meltzer, P. S. (1990) A modified CHEF system for PFGE analysis. *Technique* **2**, 36–42
16. Blonden, L. A. J., den Dunnen, J. T., Van Paassen, H. M. B., Wapenaar M. C., Grootscholten P. M., Ginjaar, H. B., Bakker, E., Pearson P. L., and van Ommen G. J. B. High resolution deletion breakpoint mapping in the DMD-gene by whole cosmid hybridization. *Nucleic Acids Res.* **17**, 5611–5621.
17. Burke, D. T., Carle, G. F., and Olson, M. V (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806–812.
18. De Jonge, P., De Jongh, F C. M , Meijers, R., Steensma, H Y., and Scheffers, W. A (1986) Orthogonal-field-alternation gel electrophoresis banding patterns of DNA from yeasts. *Yeast* **2**, 193–204.

CHAPTER 18

Cloning from Gels Following Pulse-Field Gel Electrophoresis

Peter J. Scambler and Michele Ramsay

1. Introduction

Cloning from DNA fragments fractionated by pulse-field gel electrophoresis (PFGE) offers an opportunity to isolate markers from a specific region of a genome and thus forms part of the armory of the reverse geneticist. In particular, it can be used as an adjunct to chromosome walking and jumping strategies, which are slow procedures if initiated from a single point; isolation of additional start sites facilitates the saturation cloning of a particular region of DNA. The method involves digestion of human genomic DNA with a "rare cutter" endonuclease, fractionation by PFGE, excision of the fragment of interest from the gel, and purification of DNA and its cloning into an appropriate vector (*see* Fig. 1 for summary).

There are a number of advantages and disadvantages to the technique, which need to be considered before deciding that pulse-field (PF) gel cloning is an appropriate strategy. The obvious competing technique is yeast artificial chromosome (YAC) cloning (*see* Chapter 19, this volume), though the two methods may be combined so that a particular PF fragment is cloned into a YAC vector. One advantage of PF cloning into *E. coli* vectors is that the only limit to the size of the region from which clones can be obtained is the size of DNA fragment that can be resolved by PFGE. End fragments can also be cloned from the gel, which, in conjunction with a linking library *(1),* could be used to isolate clones from adjacent restriction fragments. The DNA from the PF fragment can be cloned into plasmid, cosmid, or phage vectors; phage

From: *Methods in Molecular Biology, Vol 9: Protocols in Human Molecular Genetics*
Edited by. C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

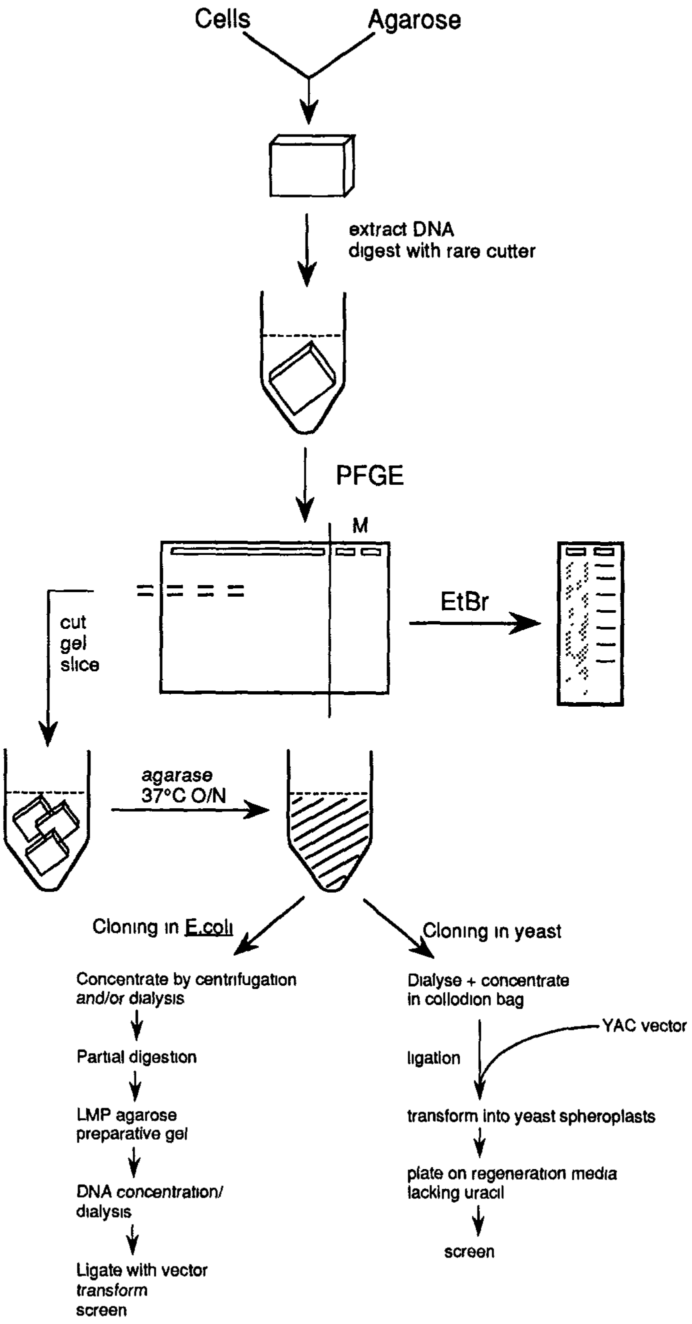


Fig. 1. Flow diagram of pulse-field cloning procedure.

vectors give the highest efficiencies. Clones obtained may generally be used immediately for mapping purposes or for further library screening. One disadvantage of the PF cloning technique is that, to be efficient, it requires a somatic-cell hybrid containing the chromosome of interest in a heterologous background. If, however, a particular PF fragment is cloned into a YAC vector and transformed into yeast spheroplasts, it is not necessary to have a somatic-cell hybrid as starting material.

1.1. Somatic-Cell Hybrids and Enrichment for the Sequence of Interest

Somatic-cell hybrids are generally exploited as the source of DNA for PF cloning, since their use considerably reduces the background of recombinants from similarly sized fragments throughout the genome. In the present discussion, it will be assumed that the investigator intends to clone a human fragment from a mouse background, though most of the steps can easily be transferred to other systems. If a choice of hybrids is available, points to consider are the complexity and copy number of the human sequences present and possible methylation drift in the region of interest. Concerning complexity, it is obvious that the smaller the region of human DNA that contains the fragment of interest, the better, since the background of human clones will be lower. Occasionally, and especially when the chromosome or subchromosomal fragment contains a selectable marker, appropriate culture conditions can increase the copy number of the human chromosome or can increase the proportion of host cells that contain the chromosome of interest. Long-term passage of cell lines may result in changes in methylation patterns, with consequent changes in the restriction fragments detected by various probes. It is therefore essential that the restriction map around the region to be cloned be checked as close to the time of the cloning experiment as possible.

Protocols for further enrichment of sequences from particular overlapping PF fragments are being developed, e.g., the "coincidence cloning" method *(2)*. Enrichment can also be enhanced by judicious choice of restriction enzyme. For instance, if the enzyme SalI detected a 1-Mbp fragment containing sequences of interest, this would be useful, since SalI fragments are generally smaller than 1 Mbp. Additional enrichment is then possible when a set of enzymes that do not cleave within the fragment of interest are used in a multiple digest of the DNA. In this case, background fragments will be cleaved and the resulting smaller fragments, under electrophoresis, will migrate further, out of the region that is to be cloned.

Published accounts of PF cloning can be found in refs. *3–5*.

2. Materials

1. PFGE apparatus. Contour-camped homogeneous field electrophoresis (CHEF) or field inversion gel electrophoresis (FIGE) are probably best suited to this technique, because they generate straight lanes of digested DNA, though orthagonal field agarose gel electrophoresis (OFAGE) has also been used successfully. Gels should be large enough to give good physical separation of DNA; some commercially available transverse alternating field electrophoresis (TAFE) apparatus is therefore unsuitable.
2. Schleicher and Schuell collodion membranes (UH 020/25) and apparatus (UH 020/2a) for concentrating and dialyzing DNA.
3. Agarase (Calbiochem or Sigma).
4. Centriprep™ columns (Amicon).
5. Elutip™ columns (Schleicher and Schuell).
6. Melting buffer (MB): 100 m*M* NaCl; 10 m*M* Tris-HCl pH 8.0; 5 m*M* EDTA.
7. Long-term gel storage buffer: 10 m*M* Tris-HCl, pH 8.0; 100 m*M* EDTA.
8. 10 m*M* Tris-HCl, 10 m*M* EDTA, pH 8.
9. TE buffer: 10 m*M* Tris-HCl, 1 m*M* EDTA, pH 8.

Details of other materials required for PFGE and YAC cloning are given in Chapters 17 and 19, this volume, respectively.

3. Methods

Essentially, two methods of cloning from PF gels will be described—the first is the cloning of small subfragments of a pulse field fragment into prokaryotic vector systems; the second, the cloning of individual large fragments in yeast. The preparation of the DNA and subsequent digestion, cloning, and transformation/transfection will be described briefly and is shown schematically in Fig. 1. Table 1 gives a short outline of the procedure.

3.1. Cloning into Prokaryotic Vector Systems

3.1.1. Preparation of DNA

Preparation of the DNA and restriction digests are carried out in low-melting point (LMP) agarose blocks as usual (Chapter 17). The blocks may contain up to 10 µg of DNA and 20–25 blocks/gel may be loaded. Digests may be done in bulk and surplus blocks stored at 4°C under 0.5*M* EDTA.

3.1.2. Preparation of the Gel

1. Routinely, 0.8% LMP agarose gels are employed for the PFGE of the DNA; gels of up to 1.5% may be used if this improves the resolution of a target fragment, but this increases the difficulty of melting the gel and removing the agarose in the later stages of the protocol.

Table 1
Summary of Standard Protocol for Cloning into Prokaryotic and Eukaryotic Vectors

Run test analytical gels to select appropriate conditions
Run a preparative gel and cut a target fragment area from the preparative zone and the flanking lanes (days 1–3)
Blot the flanking lanes and the remainder of the gel to check that the target fragment has been excised (days 3–6)
Melt the target fragment and treat it with agarase overnight (days 3,4)

A: Prokaryotic Vector Cloning	B: YAC Cloning
Concentrate DNA and run an aliquot on check gel (days 4,5)	Concentrate and dialyze DNA in S&S collodion bags; transfer to an Eppendorf tube
Partially digest and precipitate the insert DNA; ligate to vector (days 5,6)	Ligate to an appropriately digested and phosphatase-treated YAC vector; take care not to shear the DNA
Transfect/transform the ligated DNA (day 6)	Transform into yeast spheroplasts as described in Chapter 19
Screen the clone bank (day 7 on)	Screen for specific clones

2. Some electrophoresis rigs require that the gel be fixed to a plate (e.g., LKB). In these cases, it is necessary to remove some LMP agarose from over the fixing points and replace it with 1.5% normal-melting temperature agarose. This ensures that the gel remains in the correct position during electrophoresis, since LMP agarose is not very cohesive.
3. It is often useful to run one or two test gels prior to the cloning attempt, blotting the LMP gels and hybridizing them with the probe of interest. This allows an estimation of how the fragment of interest migrates under exactly those conditions that are to be used in the cloning experiments. In particular, the position of the fragment with respect to the markers should be noted; we have seen instances in which the relative migration of genomic DNA fragments and markers changes when shifting from standard to LMP agarose even when maintaining the other parameters constant. It is also important to check that the distance of migration of the target fragment from the origin is constant across the width of the gel.
4. The gel is set using a comb with a large central well and two small wells at edges.
5. The gel is loaded with markers in the outermost lanes and a single block of digested genomic DNA in the inner of the two single wells. The cen-

tral well is completely filled by inserting blocks of genomic DNA side by side. The blocks are then securely anchored by sealing the well with molten agarose.

3.1.3. Electrophoresis

1. The gel is electrophoresed under conditions known to give good resolution of the fragment to be cloned (*see* Chapter 17).
2. At the conclusion of the run, the gel is placed onto a clean tray and the two lanes at each edge (i.e., one marker lane and one genomic digest) are cut away from the rest of the gel, which is stored in 10 m*M* Tris-HCl 10 m*M* EDTA, pH 8, at 4°C.
3. The side lanes are stained in running buffer plus ethidium bromide, destained, and photographed adjacent to a ruler.
4. The distance of migration of the target fragment from the origin is now estimated from the migration distance of marker fragments and with reference to the test gels.
5. A 2- to 4-mm block of agarose around this point is now cut from the preparative gel.
6. The gap is filled with molten agarose, the edge slices replaced, and the reconstituted gel Southern blotted. The blot is hybridized with a probe recognizing the target fragment, which should demonstrate that the correct area has been excised, with the DNA in the edge lanes acting as positive control.

3.1.4. Manipulation of the Gel Slice

To lower the chances of DNA degradation and avoid complications resulting from long-term storage in EDTA, it is wise to proceed with the DNA extraction procedure immediately, rather than waiting for the hybridization result.

1. The strip of agarose containing the target fragment is washed in TE and diced in a sterile Petri dish using a sterile blade.
2. The agarose fragments are placed in a sterile container with an equal volume of MB and incubated at 65°C. The agarose takes 10–30 min to melt.
3. The melted gel is allowed to cool to 37°C and 50 U of agarase (Calbiochem) is added for each mL of gel. The agarose is digested overnight at 37°C.

3.1.5. Concentration of the DNA

1. The agarase step prevents the gel from reforming at room temperature, but if the subsequent DNA concentration steps follow directly, the solution may gel as agarose oligomers themselves become more concentrated.

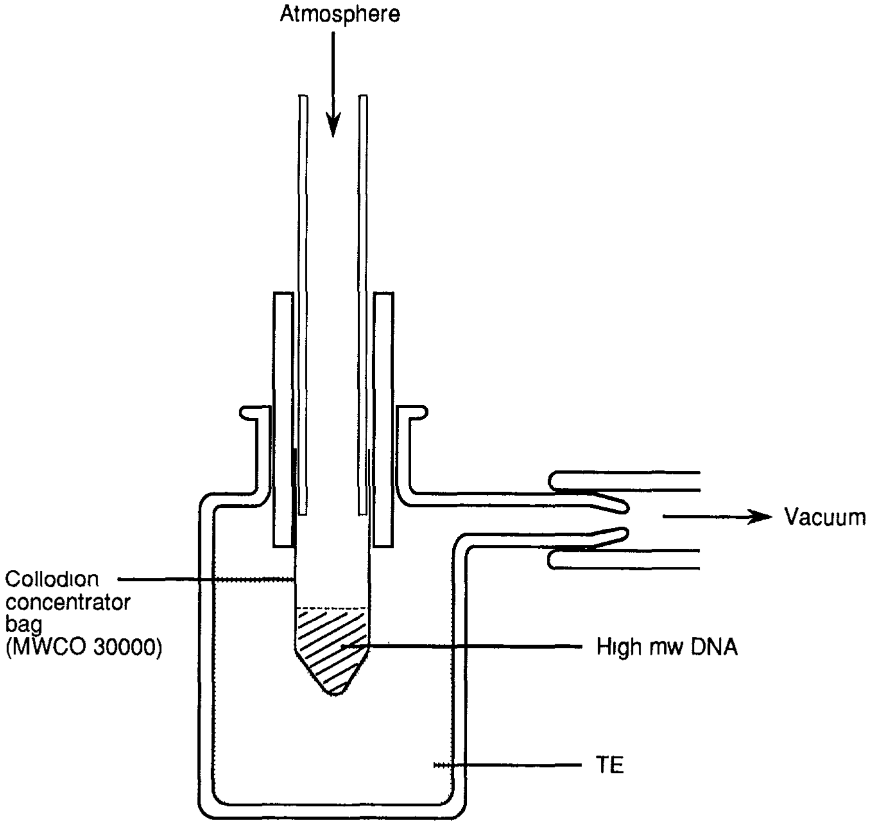


Fig. 2. Sketch of dialysis/concentration apparatus.

In order to prevent this, the overnight incubation is placed on ice for 20 min and then centrifuged at 5000*g* for 20 min at 4°C. A small pellet of gel will be obtained.

2. The supernatant is then concentrated using commercial filtrators, such as Centriprep™ (Amicon) columns, in which 15 mL can be reduced in vol to 2 mL or less, without increasing the concentration of small solutes.
3. Further concentration is achieved with a dialysis apparatus, such as that supplied by Schleicher and Schuell (Fig. 2). The sample is placed in the dialysis bag, dialyzed against TE, and concentrated by applying a vacuum above the TE. A change of TE is advisable half-way through each concentration. The final vol should be <100–500 µL.
4. An aliquot of this DNA (e.g., 5% by vol) can be run on a gel to check the integrity of DNA, which should be 50–100 kbp in length.

3.1.6. Digestion of the DNA

The DNA from the cut fragment is now prepared for cloning by digestion with an enzyme appropriate to the vector systems to be used. In order to create representative clone banks from a particular fragment, it will be necessary to conduct partial digestions in order to derive conditions that provide fragments in the required size range. For this purpose, DNA from a preparative PF gel fragment adjacent to the target fragment can be isolated in such a way as to conserve the precious DNA from the region of interest. Digested insert DNA may be ligated directly into vector, but we have found considerable selection for smaller inserts, which are more difficult to analyze subsequently. For these reasons, we have selected by size the DNA to be cloned, and this has the additional advantage of lowering substantially the risk of coligating and cloning noncontiguous fragments.

1. When appropriate reaction conditions have been determined, the insert DNA is digested and loaded into a single large well of a preparative LMP gel, with lambda/HindIII markers in flanking lanes.
2. The gel is run, and the flanking marker lanes are cut away, stained with ethidium bromide, and photographed adjacent to a rule.
3. For phage cloning, 15- to 24-kbp fragments are used for replacement vectors and the 4- to 15-kbp fragments taken for insertion vectors. Smaller fragments can be collected, but should be cloned separately from the larger fragment sample.
4. DNA is then concentrated as previously and precipitated using an Elutip™ (Schleicher and Schuell) column. The precipitate is resuspended in 10 µL of TE.

3.1.7. Cloning

1. Standard protocols for ligation and packaging are followed (*see* in this series, Vol. 4, Chapter 17), ligating 2 µL (between 20 and 200 ng) of insert with an excess (1–2 µg) of vector.
2. Following packaging and plating, the phage plaques are screened for human sequences with a BLUR (human Alu repeat) probe or total human DNA.
3. Positives are picked for secondary screening, and secondary positives are isolated for DNA preparation.
4. We commonly use the Lambda ZAP™ series as a phage-insertion vector. This allows inserts to be excised as Bluescript® plasmid recombinants during *E. coli* culture. The inserts may be cut directly from LMP gels and, following labeling and competitive hybridization with an excess of unlabeled human genomic DNA, they may be used directly to screen

cosmid libraries in order to generate a series of overlapping clones from the target fragment.

5. If end fragments are required from the PF target fragment, a proportion of the digested insert DNA should be ligated to vector digested in such a way as to produce cohesive ends for both the endonuclease used for the initial PF gel digestion and the partial digestion. The Charon series of phages are generally suitable for this purpose (6). For a specific example using EMBL3-derived phage, *see* ref. *3*.

3.1.8. Modifications

1. A rapid test using dot-blot apparatus to show that the correct gel slice has been cut in preparation for cloning:

   Locate the target fragment as before, but continue with the DNA extraction procedure directly, rather than waiting for a Southern blot hybridization result. A hybridization result can be obtained more quickly by taking a couple of 2-mm-wide strips of agarose spanning the area of interest, treating with agarase, taking a small sample, melting it by heating at 65°C, and denaturing the DNA by bringing the sample to 0.5*M* NaOH. The samples are then dot-blotted using a commercial manifold and hybridized to determine the precise location of the fragment of interest within the preparative gel.
2. To increase the yield of small inserts using a polymerase chain reaction (PCR) method:

   The small (<4kbp) fraction of digested insert DNA is ligated to a vector, such as M13, and the ligated DNA is precipitated and resuspended in TE. Oligonucleotides complementary to vector sequences flanking the polylinker are then used to amplify the insert sequences. The amplified DNA can then be digested with the polylinker enzyme(s) of choice and cloned into another vector. *See* ref. *7* for further details of this technique. Alu-PCR *(8)* may also be employed.

3.1.9. Assessment of Clones

In order to assess the success of a particular PF cloning experiment, it is necessary to determine the total number of human recombinants obtained, the number of duplicate clones (we have noted a surprisingly large number of identical clones generated by separate cloning experiments), and the proportion of recombinants that originate from the target fragment. It is difficult to calculate the enrichment factor if the copy number of the target fragment within the hybrid is unknown or if the proportion of cells carrying the human chromosome is uncertain. One quick method of assessment is to compare the number of clones obtained from the target fragment, as a per-

centage of the total number of human clones analyzed, with the percentage of clones expected to map to the target fragment by chance, following the analysis of the same number of human recombinants from a conventional library constructed from the somatic-cell hybrid.

The clones should be mapped back to the target region, using PFGE. It is not sufficient to show that they hybridize to a fragment of the same size as the target fragment; they should detect restriction fragments with other enzymes in such a way that they are compatible with a predetermined map. Occasionally, partial methylation or consistent partial digestions can give a diagnostic pattern of fragments; probes that detect the target fragment then also detect additional fragments *(4)*.

3.2. Cloning of PF Fragments into YAC Vectors

An analysis of the region under study by PFGE may identify a particular fragment that would be useful to have cloned and propagated in a heterologous background; YAC cloning vectors are most convenient for this purpose. This strategy is most suitable when the target fragment almost certainly contains a gene of interest. It gives one greater control over the insert within the resulting clone than would be obtained when screening a YAC library generated by EcoRI partial digests. Further applications would be in cloning a large gene in its entirety in preparation for expression studies, in cloning genes containing certain mutations, in cloning chromosomal breakpoints, and in cloning regions that are purported to be unstable in prokaryotic systems. Once the desired recombinant has been isolated, the insert can easily be subcloned into phage or plasmid vectors if necessary. A more comprehensive procedure for YAC cloning is described in Chapter 19. Variations will be described briefly.

The main disadvantage of the method is that the larger the fragment, the more difficult it will be to clone, mainly because of the problem of mechanical shear of the DNA. It is also not particularly suitable when the closest markers are thought to be further than 0.5 Mbp from the region of interest. In that case, it would be more efficient to clone the end of a PF fragment as discussed above and to use the resultant clone in further mapping and walking.

3.2.1. Source of DNA

In this case, it is not important to use somatic-cell hybrids as starting material, because a unique sequence probe will be used in screening for the single correct recombinant. Lymphocytes can therefore be used as the source DNA. The resulting YAC clone will contain the entire region of interest in a yeast background. As before, it is essential to check that the target fragment is present in the donor DNA and to assess its migration under the conditions that will be used for the preparative gel.

1. The DNA is digested with the suitable rare cutting restriction enzyme, electrophoresed in LMP agarose, and cut from the gel as described above.
2. The gel slice is liquefied with agarase (50 U/mL in MB), and care is taken to minimize mechanical shear of the DNA (stirring gently in solutions, limiting pipeting steps, and using tips with their ends cut off with a sterile blade for all manipulations).
3. The liquefied DNA is placed in a collodion bag and dialyzed and concentrated with two changes of TE. The volume is reduced to 100 µL and transferred to an Eppendorf tube.

3.2.2. Cloning and Transformation

1. The YAC vector is prepared as described in Chapter 19, and ligations are done in the standard way.
2. A small aliquot (5 µL) of the ligation is used with 5 µg of carrier DNA (usually herring-sperm DNA) for transformation into yeast spheroplasts. The carrier DNA significantly increases the number of transformants.
3. The resulting clones are plated onto selective media either in 2.5% agar or in alginate.

3.2.3. Screening

It is essential to use a rapid screening procedure, since a single clone is sought from a background of clones with similarly sized inserts. Picking individual transformants is therefore undesirable. Essentially, two options exist:

1. The transformants can be plated in 2.5% agar and a pin-cushion device used to transfer them onto the surface of the plate.
2. They can be plated in alginate in such a way that the recombinants grow on the surface of the plate after the cells have regenerated their cell walls within the alginate.

Both methods result in clones on the surface of a plate, from which they may be transferred to nylon membranes, and the DNA extracted and screened with the appropriate unique-sequence probe by standard methods described in the YAC-cloning chapter (Chapter 19).

3.2.4. Clone Assessment

In this case, a single recombinant of known size is selected, and its integrity can be verified by checking for the appropriate rare cutter restriction sites predicted by the PF map. It can also be digested with a variety of frequently cutting restriction enzymes and electrophoresed in parallel with genomic DNA, and the blot hybridized with the probe used for screening.

Having obtained the correct clone, it can be investigated for breakpoints or specific mutations, or it can be tested for the presence of coding sequences by one or more of the following means:

1. Hybridization of the YAC to zoo blots and Northern blots of RNA, pre-annealing repeated sequences to an excess of unlabeled genomic DNA prior to filter-hybridization.
2. Screening the YAC with RNA from a suitable tissue;
3. Checking for the presence of HTF islands; and
4. Subcloning the YAC and using single copy probes to hybridize to zoo blots, Northern blots, and cDNA libraries.

4. Notes

The following indicates possible problems and remedies:

1. Agarose is not melting:
   a. Make sure that the blocks are equilibrated to a low (5 m*M*) EDTA concentration.
   b. Try mixing LMP agarose with ultralow-gelling temperature agarose, e.g., Sea Prep from FMC Bioproducts.
   c. *In extremis,* agarose fragments can be sheared using a 21-gage needle. The DNA is usually still of sufficient length to give reasonable cloning efficiencies in insertion phage vectors.
2. Agarose sets during DNA concentration:
   a. Check that melting is complete.
   b. Increase the concentration of agarase in the overnight digestion stage.
   c. Add an extra 4°C cooling/centrifugation step.
3. The insert DNA is too small for the desired cloning system:
   a. Prepare new buffers.
   b. Increase the concentration of EDTA in gel-storage solutions; be sure to rinse well before attempting to melt the gel.
   c. Add prewarmed EDTA to the overnight agarase step, to a final concentration of 100 m*M*.
   d. Cut the ends from pipet tips to give wider bores and reduce hydrodynamic shear.
4. Low yields of recombinant clones:
   a. Run controls for ligation, packaging, and transformation.
   b. Check that the size of the insert DNA is suitable for the vector chosen.

   c. Check that traces of agarose or contaminants from the agarose are not inhibiting enzyme reactions; change the batch of agarose.
   d. Consider PCR cloning (Section 3.1.8).
5. Low enrichment for fragment of interest:
   a. Check the gel electrophoresis. Is the target fragment migrating the same distance from the origin throughout the gel? If not, check the electrode configuration; check the buffer or coolant circulation (is there a temperature gradient across the gel?); check that the gel tank is level; check that the gel is of uniform thickness; if the apparatus is similar to the Biorad rig (in which the gel is not fixed in place), check that one edge of the gel is not lifting as a result of uneven buffer flow or excess buffer in the tank.
   b. Cut a narrower strip of gel around the target fragment.
   c. Consider multiple digest protocols as described above.
6. Coligation of insert fragments:
   a. Increase the vector concentration in ligation.
   b. Try to obtain better size selection of insert fragments; check that the type of vector you are using is appropriate to the inserts you are attempting to clone.
   c. Some vector-cloning systems will allow you to treat the insert with phosphatase before cloning. Do this prior to the precipitation of DNA. Phosphatase treatment will decrease ligation efficiency to some extent.
   d. This may not be a great problem, provided that it has been detected and appropriate steps are taken to check the map position of each subclone obtained from a recombinant.

References

1 Collins, F. S (1988) Chromosome jumping. *An Analysis: A Practical Approach.* IRL, Oxford, UK, pp. 73–93.

2. Marchuck, D., Cole, J., Cantor, C , Weissman, S., and Collins, F. (1988) Coincidence cloning: a method for selective cloning of sequences shared between DNA samples. *Am. J. Hum. Genet.* **43,** A773.

3. Michiels, F., Burmeister, M., and Lehrach, H. (1987) Derivation of clones close to MET by preparative pulsed field gel electrophoresis. *Science* **236,** 1305–1308.

4 Anand, R., Honeycombe, J., Whittaker, P. A., Elder, J. K., and Southern E. M (1988) Clones from an 840-kb fragment containing the 5' region of the DMD locus enriched by pulsed field gel electrophoresis. *Genomics* **3,** 177–186

5. Ramsay, M., Wainwright, B. J., Farrall, M., Estivill, X., Sutherland, H., Ho, M-F., Davies, R., Halford, S., Tata, F., Wicking, C., Lench, N., Bauer, I., Ferec, C , Farndon, P.,

Kruyer, H., Stanier, P., Williamson, R., and Scambler, P. J. (1990) A new polymorphic locus, D7S411, isolated by cloning from preparative pulse-field gels is close to the mutation causing cystic fibrosis. *Genomics* **6**, 34–47.

6. Dunn, I. S. and Blattner, F. R. (1987) Charons 36-40: multi enzyme, high capacity, recombination deficient replacement vectors with polylinkers and polystuffers. *Nucleic Acids Res.* **15**, 2677–2697.
7. Ludecke, H-J., Senger, G., Claussen, U., and Horsthempke, B. (1989) Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. *Nature* **338**, 348–350.
8. Nelson, D. L., Ledbetter, S. A., Corbo, L., Victoria, M. F., Ramirez-Solis, R., Webster, T. D., Ledbetter, D., and Caskey, C. T (1989) Alu polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proc Natl. Acad Sci. USA* **86**, 6686–6690.

CHAPTER 19

Yeast Artificial-Chromosome (YAC) Cloning Systems

Michele Ramsay and Carol Wicking

1. Introduction

Yeast artificial-chromosome (YAC) cloning systems are used to clone large contiguous segments of DNA from any organism into suitable vectors in such a way that the recombinants can be transformed into yeast *(Saccharomyces cerevisiae)* cells, where they are stably propagated. The basic strategy for producing artificial chromosomes in yeast was first described in 1983 by Murray and Szostak *(1)*. In 1987, Burke et al. *(2)* demonstrated that artificial chromosomes in yeast were useful for cloning very large segments of exogenous DNA.

The cloning vectors contain sequences necessary for an autonomously replicating, stable chromosome with suitable selectable markers in yeast (Fig. 1). The donor DNA consists of very high-mol-wt DNA, which is isolated either in liquid or in low-melting-point (LMP) agarose blocks, digested into suitable lengths with restriction enzymes, and size-selected either on sucrose gradients or by focusing using pulse-field gel electrophoresis (PFGE). After ligation, the recombinants are transformed into spheroplasts of a yeast strain auxotrophic for bioselectable markers and are plated onto the appropriate selective medium. Subsequent screening of the YAC recombinants can be performed by yeast-colony hybridization *(2)* or by polymerase chain reaction (PCR) *(3)*. A general scheme for YAC cloning based on the construction of an EcoRI partial digestion library is given Fig. 2.

In this review, the advantages, applications, and general strategies of YAC cloning are discussed briefly. A set of basic methods for the production, screen-

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

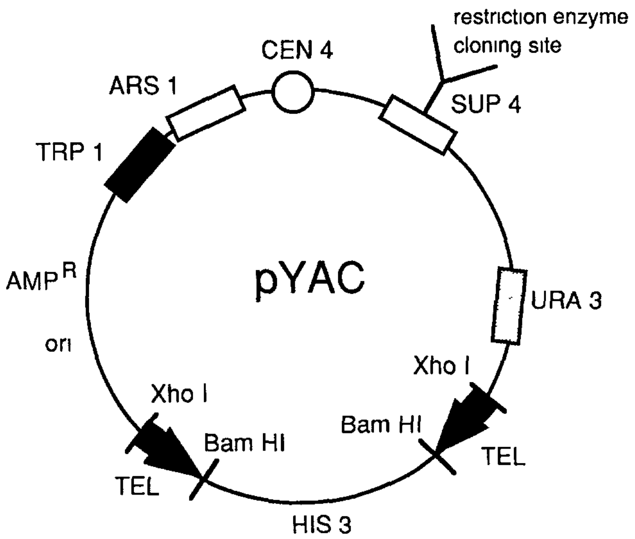


Fig 1. Schematic representation of a typical YAC series vector (adapted from ref. *34*). It consists of two telomere repeat sequences that originate from *Tetrahymena* and are separated by a BamHI "stuffer" fragment, which is cleaved from the vector prior to ligation. The cloning site separates the vector into two arms, each of which contains a selectable marker in yeast and one of the telomere sequences. One arm contains *TRP 1* as a yeast-selectable marker, the sequences for replication in yeast *(ARS 1)*, and an *S. cerevisiae* centromere, as well as the origin of replication for *E. coli* and an ampicillin-resistant marker for selection during the preparation of the vector. The other arm contains the *URA 3* selectable marker in yeast. The cloning site is within the *SUP 4* gene, which, if introduced intact into AB1380, suppresses a mutation at the ade2 locus in the host, resulting in a color change from red to white in the presence of limiting concentrations of adenine. This provides a convenient phenotypic selection for recombinants. If interruption of the *SUP 4* gene occurs, recombinants will grow red and nonrecombinants with an intact suppressor will grow white *(35)*. Several modifications and variations to the original vectors have been made and these will be discussed in the section on choice of vector.

ing, and initial characterization of YACs will follow, and alternative ways of achieving the same results will be mentioned. This technology goes hand in hand with that of PFGE and/or field-inversion gel electrophoresis (FIGE; *see* Chapters 17 and 18, this volume). Long-range maps using infrequently cutting restriction enzymes are an important component for the characterization and correct positioning of YAC clones.

1.1. Advantages and Applications

The YAC cloning system has been successfully applied in the cloning of insect, mammalian, and plant DNA *(4–9)*. The greatest advantage of YAC

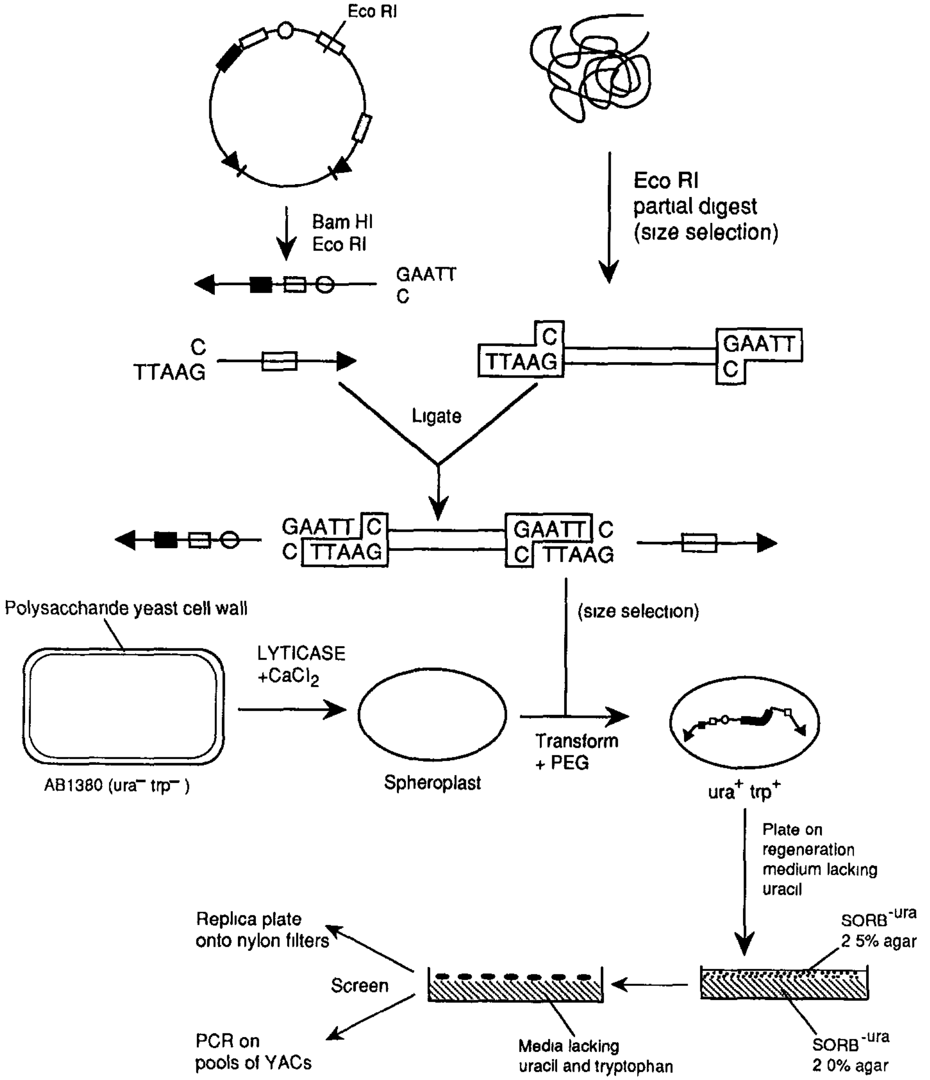


Fig. 2. General scheme for YAC cloning based on the construction of an EcoRI partial-digestion library in pYAC4. DNA is digested, ligated, and transformed into yeast spheroplasts in the presence of PEG. Size selection steps can be performed before ligation, after ligation, or both. Following transformation, colonies are regenerated within agar lacking uracil and subsequently transferred to the surface of a plate lacking both uracil and tryptophan. Colonies growing red on this plate should be transformants. Note that colonies may not grow red when embedded in agar, since the synthesis of the red pigment requires oxygen, which will be limiting in the case of those colonies deep in the agar. Transformants are replicated and screened by the desired method (*see* Section 3.6).

cloning is that it allows the cloning of far larger DNA fragments than those accommodated by existing prokaryotic cloning systems. YAC clones are stable in the size range of 50 kilobase pairs (kbp) to over 1 megabase pairs (Mbp), although in practice there is a strong bias for cloning smaller fragments over large ones.

Because of this increased size capacity, YACs can be used to clone complete genomic gene sequences in cases in which the genes are too large to be cloned in prokaryotic systems, for example, the Duchenne muscular dystrophy (DMD), Factor VIII, and cystic fibrosis (CF) genes *(10–12)*. They can also be used in the cloning of candidate genes for diseases in which probes that are linked to the disorder have been identified. If the markers are tightly linked to the gene of interest, positive YAC clones may be examined for coding sequences, or, if flanking markers have been identified, a pulse-field map could be constructed to link them physically, and specific pulse-field fragments could be targeted for cloning into YAC vectors (*see* Chapter 18). When markers are at an uncertain distance from the gene of interest, end fragments of the YAC can be subcloned and used for rapid directional walking along the chromosome.

In the case of YAC cloning, the host organism is eukaryotic, and sequences that have been reported to be unstable, underrepresented, or absent in libraries constructed in prokaryotes may well be clonable in this system. Stability of these sequences may also be conferred by the simultaneous cloning of adjacent DNA sequences. Methods for transferring YACs into mammalian cells using cell fusion, electroporation or microinjection are being explored *(13)*, and YACs that are known to contain genes could then be used for expression studies in eukaryotic systems and for the production of animal models of human disease.

Furthermore, YAC technology is being used in general mapping projects in which the aim is to clone entire chromosomes or genomes, and a very elegant study was recently described in which YACs were used to bridge the gaps between cosmid and phage contigs *(5)*.

2. Materials

Most of the equipment for YAC cloning will be present in a standard laboratory that is set up for work with recombinant DNA. Incubators, both shaking and stationary, are needed at 30°C. Standard equipment for PFGE and FIGE are needed, though the latter is not essential. Mechanical replicating devices and pin-cushion replicators may soon be available commercially, though at present they are constructed "in house."

2.1. Vector Preparation

1. BamHI or other appropriate restriction enzyme.
2. Calf-intestinal phosphatase (Boehringer Mannheim).

2.2. Donor DNA

1. Materials for isolation of DNA in agarose blocks (*see* Chapter 17).
2. Appropriate restriction enzyme (e.g., EcoRI for cloning partial digests into pYAC 4).
3. Lambda DNA concatamers (mol-wt markers; Pharmacia).
4. Agarase (Calbiochem).
5. Collodion bags (UHO20/25, mol-wt cutoff, 25,000) and concentrating apparatus (Schleicher & Schuell).

2.3. Ligation

1. T4 DNA ligase.

2.4. Transformation

1. YPD (1 L): 10 g of bacto-yeast extract, 20 g of bacto-peptone, 20 g of glucose. Add distilled water to make 1 L, adjust pH to 5.8, and autoclave. For plates, add 20 g of bacto-agar prior to autoclaving.
2. SORB without uracil (1 L): 182 g of sorbitol, 7.6 g of yeast nitrogen base without amino acids (Difco), 600 mg of amino acid mix without uracil, 20 g of glucose. Add distilled water, adjust pH to 5.8, and autoclave. For plates, add 17 g of agar prior to autoclaving; for regeneration TOP media, add 25 g of agar.
   Note: For the above media, some people prefer to filter-sterilize the glucose as a 50% solution and to add 40 mL to 1 L of sterile media.
3. Amino acid mix: Weigh solids together and mix in a blender. The quantity given is for 100 L of medium, and the final concentration is given. Omit appropriate amino acids to produce mixes for single- and double-selection media and use the adjusted quantity of solids per liter of medium (e.g., 600 mg/L for –ura medium and 560 mg/L for medium lacking uracil and tryptophan).

Compound	g	Final concentration, µg/mL
arginine-HCl	4	40
histidine-HCl	2	20
isoleucine	6	60
leucine	6	60

lysine-HCl	5	50
methionine	2	20
phenylalanine	5	50
threonine	20	200
tryptophan	4	40
tyrosine	5	50
adenine hemisulfate	1	10
uracil	2	20

4. SCE (1 L): 182 g of sorbitol; 100 mL of 1*M* sodium citrate, pH 5.8; 100 mL of 100 m*M* EDTA. Autoclave. For SCEM, add β-mercaptoethanol to a final concentration of 30 m*M*.
5. STC (1 L): 182 g of sorbitol; 10 mL of 1*M* Tris, pH 8.0; 10 mL of 1*M* $CaCl_2$. Autoclave.
6. Polyethelene glycol (PEG) (100 mL): 20 g of PEG 8000; 1 mL of 1*M* Tris, pH 7.5, 1 mL of $CaCl_2$.
   Note: PEG solution may be autoclaved without the $CaCl_2$, which can be added prior to each use.
7. SOS (100 mL): 75 mL of 1*M* sorbitol, 25 mL of YPD, 700 µL of 1*M* $CaCl_2$, 270 µL of 1% uracil, 4.7 g of sorbitol. Filter-sterilize and store in small aliquots.
8. Sorbitol, 1*M* (Sigma or Fisher).
9. Lyticase (No. L-8137, Sigma).
10. Carrier DNA (from calf thymus or herring sperm).

2.5. Characterization of Library

1. Lyticase: 2 mg/mL in SCE.
2. LMP agarose: 1% in 1*M* sorbitol.
3. Proteinase K: 20 mg/mL in 0.25*M* EDTA, 1% sarkosyl.
4. Phenylmethylsulfonyl fluoride (PMSF): dissolve at 40 mg/mL in isopropanol or ethanol; dilute to 0.04 mg/mL in TE for use.

2.6. Screening

1. AHC^- (1 L) (rich medium lacking uracil and tryptophan): 5.6 g of yeast nitrogen base without amino acids, 10 g of casein hydrolysate acid, low salt (US Biochemical Corporation Catalog no. A-9126), 20 mg of adenine hemisulfate, 20 g of glucose. Add distilled water and adjust the pH to 5.8. Autoclave. For plates, add 20 g of agar prior to autoclaving.
2. SD^- (1 L) (medium lacking uracil and tryptophan): 6.7 g of yeast nitrogen base without amino acids, 560 mg of amino acid mixture lacking

uracil and tryptophan, 20 g of glucose. Add water and adjust the pH to 5.8. Autoclave. For plates, add 20 g of agar prior to autoclaving.
3. Microtiter plates, 96-well (Nunc).
4. Nylon filters (Hybond-N, although Hybond-N^+ may give a stronger hybridization signal and is recommended if the filters are to be used many times; Amersham).
5. Zymolyase-100T (Seikagaku Kogyo, Tokyo; available from ICN).
6. Denaturing solution: 1.5*M* NaCl, 0.5*M* NaOH.
7. Neutralizing solution: 1.5*M* NaCl, 1*M* Tris, pH 7.4.
8. Ribonuclease A: 10 mg/mL stock solution heated to 100°C for 15 min and stored frozen.

3. Methods

There are several different strategies for the construction of YAC libraries, and choices exist at each step along the way. These strategies will be described with suitable references, followed by details of methods.

3.1. Vector

3.1.1. Choice of Suitable Vector

The pYAC series includes vectors that allow cloning of inserts generated by EcoRI (pYAC4), NotI (pYAC5), or blunt-end (pYAC3) enzyme digestions *(2)* (Fig. 1). These vectors have been further modified to provide additional rare cutter cloning sites (pYAC-RC, *[14]*), a mammalian selectable marker (pYAC4-*neo, [15]*), and *Drosophila* P-element ends with a suitable selectable marker *(4)*. In addition, specific derivatives of these vectors allow cloning of telomeres with their adjacent DNA segments *(16–18)* and others allow the transfer of clones from *S. cerevisiae* to *S. pombe* and have facilitated the cloning of *S. pombe* centromeres *(19)*.

Another set of YAC vectors have been described by McCormick et al. *(20)*, and they have been modified into a set of yeast-fragmentation vectors that can be used to fragment YACs at Alu repeat sequences *(21)*. In most cases, the YAC vectors are available upon request from the laboratory responsible for their production.

3.1.2. Preparation of Vector (see *Note 4)*

Owing to the presence of a bacterial origin of replication along with an ampicillin-resistant gene, pYAC vectors can be propagated in *E. coli* and DNA isolated by standard methods *(22)*. The vectors are digested with BamHI (to generate sequences that will heal into functional telomeres) and the enzyme that defines the cloning site in the vector of choice. This should yield three

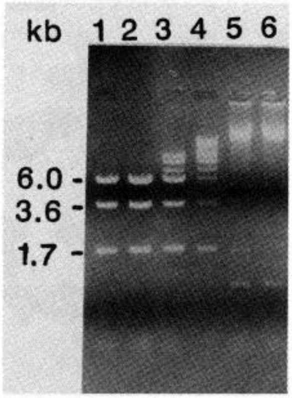


Fig. 3. Digestion and phosphatase treatment of pYAC4. Following digestion with BamHI and EcoRI, the vector was treated with increasing concentrations of calf-intestinal phosphatase and self-ligated to check the efficiency of phosphatase reaction. Lane 1, 0.1 U phosphatase/μg vector; Lane 2, 0.05 U/μg; Lane 3, 0.01 U/μg; Lane 4, 0.005 U/μg; Lane 5, 0.001 U/μg; Lane 6, no enzyme. For scaleup purposes, the vector was treated with 0.05 U/μg phosphatase.

DNA fragments: a left chromosome arm, which includes the centromere (6.0 kbp); a right chromosome arm (3.6 kbp); and the "stuffer" fragment, which separates the telomere sequences in the circular plasmid (1.7 kbp) (Fig. 1). An aliquot of the digested vector can be checked by conventional agarose gel electrophoresis. If digestion is complete, the vector should be treated with calf-intestinal phosphatase (although bacterial alkaline phosphatase has also been used successfully) to prevent self-ligation reactions. It is desirable to titrate the phosphatase and choose the concentration that successfully eliminates self-ligation of the vector, as judged by agarose gel electrophoresis. Figure 3 shows the digested YAC vector and a typical titration of phosphatase activity followed by self-ligation reactions.

3.2. Donor DNA

3.2.1. Choice of Source DNA

As with conventional libraries, the source of donor DNA is dependent on the final application. For the preparation of genomic libraries, DNA can be extracted from fresh lymphocytes, transformed lymphocyte cell lines, or any other suitable source, such as fresh mouse spleen. Chromosome-specific libraries can be prepared from somatic cell hybrid DNA or from flow-sorted chromosomes, and single pulse-field fragments can be cloned from a LMP agarose slice if a PF map of the area is available. The latter method is described in Chapter 18.

3.2.2. Preparation of DNA (see Notes 3 and 7)

Regardless of the source, it is essential to prepare the DNA in such a way that it remains of very high mol wt. When the technique was first described, DNA was isolated in liquid (taking great care not to shear the DNA) and size-fractionated on sucrose gradients *(2)*. Subsequently, donor DNA has been extracted and size-fractionated in LMP agarose using PFGE *(20,23)*. Though both methods work well, the latter requires less starting DNA and minimizes DNA shearing caused by handling. A suitable strategy for the isolation of high-mol-wt DNA in LMP agarose blocks is given in Chapter 17 on PFGE. In this laboratory, each block is prepared to contain approx 10 µg of DNA (1×10^6 cells/80 µL block), although three to four times this concentration has been used successfully *(20)*. For a method for the isolation of high-mol-wt DNA in liquid, the reader is referred to *(24)*.

3.2.3. Digestion of DNA

High-mol-wt DNA is digested with the appropriate enzyme, either partially or to completion, depending on the vector chosen. A sample should be checked for satisfactory digestion by analysis on a PFG before proceeding. In the case of partial digestion, conditions should be established on a small scale and subsequently scaled up for preparative purposes. A protocol for partial digestion of DNA with EcoRI in preparation for ligation into pYAC 4 is outlined below. In this case, the DNA is prepared and digested in LMP agarose. Figure 4A shows a small-scale titration of enzyme to determine appropriate conditions, and indicates the samples of choice for scale-up purposes. It is also possible to generate satisfactory partial digestions by keeping the enzyme concentration constant and varying the time of digestion, though the enzyme activity may decrease with time.

3.2.3.1. Partial EcoRI Digestion of DNA in Agarose Blocks

1. Equilibrate the blocks overnight in TE (10 m*M* Tris-HCl, 1 m*M* EDTA, pH 8.0).
2. Titrate the enzyme in such a way that a range of concentrations spanning 2–0.002 U/µg DNA are prepared in a final vol of 120 µL (the final vol of digest will be 200 µL, with 80 µL contributed by the block). Prepare enzyme dilutions in a reaction mixture containing 1× EcoRI restriction buffer, 0.1 mg/mL DNase-free BSA, and 0.01*M* spermidine. Keep all solutions on ice.
3. Add an equilibrated block to each digest and leave on ice for 30 min.
4. Place at 37°C for 1 h.
5. Stop the reaction by adding EDTA (to 0.05*M*); place on ice immediately.

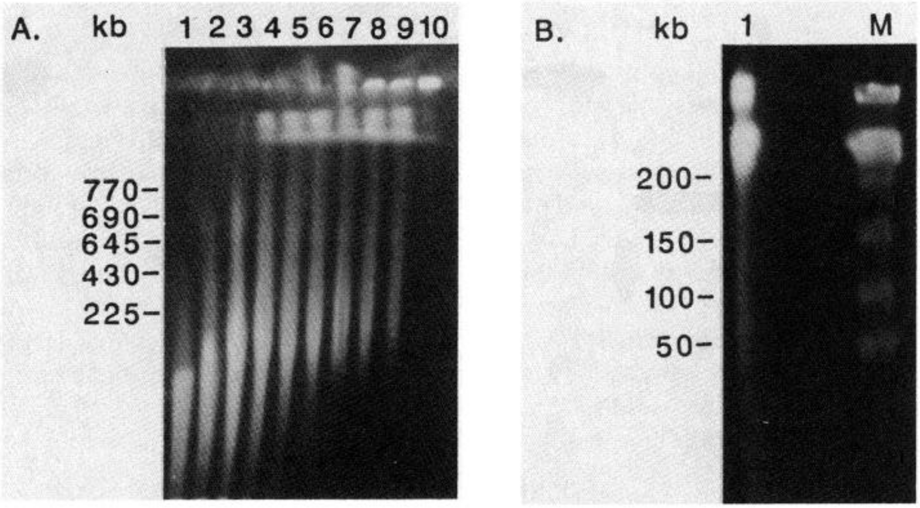


Fig. 4. (A) EcoRI partial digests using varying enzyme concentrations. Lane 1, 20 U/10 µg block; Lane 2, 10 U/block; Lane 3, 5 U/block; Lane 4, 2.5 U/block; Lane 5, 1.25 U/block; Lane 6, 0.63 U/block; Lane 7, 0.31 U/block; Lane 8, 0.16 U/block; Lane 9, 0.08 U/block; Lane 10, no enzyme. For scaleup purposes, 0.08 and 0.16 U/block were chosen. (B) Focusing of DNA above 200 kbp into the compression zone of a 0.8% LMP agarose PFG. Lane 1, high-mol-wt DNA sample; M, λ concatamers. Gel was run for 8.5 h, 200 V, 10-s pulse, 0.5× TBE in a Bio-Rad CHEF apparatus.

3.2.4. Size Fractionation of DNA

The size-fractionation step, which follows appropriate enzyme digestion, is considered essential to remove small DNA fragments that will transform preferentially in this system. This step may be performed prior to ligation, but, if included after ligation, may aid in the elimination of vector from ligation products. Two size-selection steps, both before and after ligation, may be performed. The limit one chooses for size selection of DNA depends on what is considered to be the optimal size for the resultant clones; usually one aims to get them as large as possible. In this laboratory, a convenient cutoff for general library construction is considered to be 200–250 kbp. Any attempt to increase significantly the average insert size may result in low transformation efficiencies. A general strategy for size fractionation of DNA in LMP agarose is given below. For corresponding methods of DNA preparation in liquid, the reader is referred to *(24)*. Size fractionation in agarose is conveniently achieved by focusing all DNA above a chosen size limit into the compression zone of a PFG. This band can then be cut from the gel and the DNA prepared for cloning by treatment with agarase and subsequent dialysis and concentration in collodion bags *(23)*, or the LMP gel may be melted at 65°C in

25 m*M* NaCl and used directly *(20)*. Figure 4B shows size fractionation achieved by focusing all DNA above 200 kbp into the compression zone of a LMP PFG, as judged by simultaneous electrophoresis of λ concatamers as size markers. Electrophoresis conditions will vary according to the gel apparatus used and the desired size of the YAC clones. Conditions for this gel are presented only as a guideline, and appropriate parameters should be established for each system prior to fractionation of a prepared DNA sample.

3.2.4.1. Size Fractionation of DNA by Focusing in an LMP Agarose PFG

1. Load several blocks of DNA into one large well of a 0.8% LMP agarose PFG prepared in 0.5× TBE (89 m*M* Tris-borate, 1 m*M* EDTA, pH 8.3). DNA will either be present as blocks of digested DNA, or, if the blocks have been melted for ligation and then reset, it will be necessary to remelt the DNA and load it as a liquid into the dry well of the gel. Load appropriate mol wt markers (λ concatamers) into adjacent wells on either side of the sample.
2. Focus DNA of the desired mol-wt range into the compression zone of the gel, using predetermined conditions.
3. Cut off the marker tracks and a small section of sample DNA track and stain them with ethidium bromide (Fig. 4B). Examine stained segments of gel under UV light and, using a scalpel, make nicks in the gel corresponding to the compression zone.
4. Align stained segments with the remainder of the gel, which has been stored at 4°C during staining. Cut out the appropriate region of the unstained gel, along with the high-mol-wt DNA that has remained in the wells.
5. Equilibrate the gel in an excess of 100 m*M* NaCl, 10 m*M* EDTA for 1 h on ice. Remove buffer and add fresh buffer at a vol equal to the vol of the gel.
6. Place at 68°C until the gel has melted (10–30 min). Cool to 37°C and add 40 U of agarase/mL of solution, stirring gently with a pipet tip to facilitate mixing. Leave at 37°C for 4 h or overnight.
7. If necessary, concentrate the DNA in a collodion bag, as described in Chapter 18.

3.3. Ligation (see *Note 8)*

No special consideration is required for ligation of YAC vectors to insert DNA. Digested, phosphatased vector is generally mixed with an equal weight of insert DNA and incubated overnight at 15°C in the presence of standard buffer and T4 DNA ligase. When DNA has been extracted in LMP agarose

that has not already been liquefied by agarase treatment, it is advisable to melt the agarose prior to ligation (agarose will solidify on incubation at 15°C). Small-scale ligations can be performed for initial characterization of DNA preparations, but then it is advisable to scale up the ligation to include sufficient potential transformants for complete library construction. Ligations are stable over long-term storage at 4°C. Burke and Olson *(24)* recommend phenol/chloroform extraction of ligations prior to storage in order to eliminate exonuclease activity, but we have not found this to be necessary.

3.4. Transformation

The most efficient method of introducing DNA into yeast cells is by PEG-mediated spheroplast transformation *(25)*. This involves enzymatic elimination of the yeast cell wall and transfer of DNA across the spheroplast membrane in the presence of PEG. One disadvantage of this procedure is that the transformed spheroplasts require at least one round of replication embedded in agar. It is only after regeneration of their cell wall that yeast cells are able to replicate on the surface of an agar plate. This poses problems for subsequent handling and screening, since each clone must somehow be removed from within the agar and grown on the surface to allow standard replica-plating and screening procedures (*see* Section 3.6). Attempts to avoid this step have included plating transformed spheroplasts in a thin layer of alginate in such a way that the colonies burst through and replicate on the surface of the plate *(26,27;* with further modification, *28)*. Alternatively, transformation using the lithium acetate procedure *(29)* does not require spheroplast formation and therefore circumvents the need to regenerate in agar. However, efficiencies reported with this procedure are far lower than those achievable with the spheroplast-transformation method of Burgers and Percival *(25)*. The latter procedure has remained the method of choice for YAC cloning, and a step-by-step summary is given below. It is advisable to include a control for transformation efficiency using the yeast cloning vector YCP50, which should give between 10^5 and 10^6 colonies/μg. A good high-mol-wt ligation will yield 1000 colonies/μg, but efficiencies of 200–300/μg are acceptable.

It is worth noting that one group has reported that inclusion of spermine and spermidine during the transformation procedure greatly alleviates the bias against transforming larger fragments of DNA, at least in the case of agarose-purified DNA *(20)*.

Figure 5 shows the growth of yeast transformants embedded in agar.

3.4.1. Preparation and Transformation of Yeast Spheroplasts

1. Grow AB1380 overnight in liquid culture with good aeration at 30°C in 50 mL of YPD to an OD_{660} of 2–4.
2. Pellet the cells at 500*g* for 5 min.

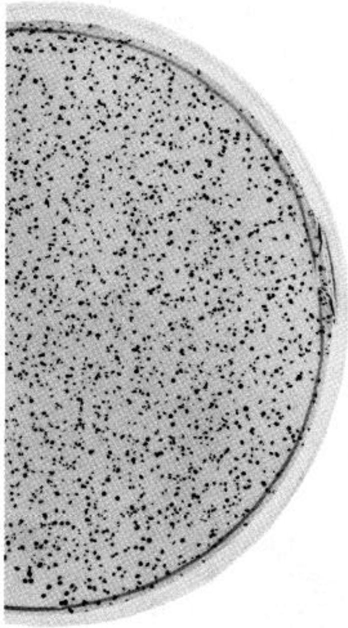


Fig. 5. Regeneration of YAC clones within a layer of 2.5% agar. Colonies grow embedded at different depths and orientations in the agar.

3. Resuspend the cells by vortexing in 25 mL of sterile distilled water and pellet again (500*g*, 5 min).
4. Resuspend the cells in 25 mL of 1*M* sorbitol and pellet again (500*g*, 5 min).
5. Resuspend the cells in 15 mL of SCEM. Dilute a sample 1/10 in distilled water and read the OD_{660} as a baseline value. For spheroplast preparation add 2000 U of lyticase to the 15-mL cell suspension and incubate at 30°C. Remove samples and measure the OD_{660} of a 1/10 dilution after 10 min and again every 5 min until the OD has fallen to 20–30% of the baseline value. Spheroplasting can also be monitored by microscopy, by mixing a sample with 1% SDS and checking for lysis by the appearance of "ghosts." Optimally, sheroplasting will proceed to the desired level in 20–30 min.
6. Pellet the spheroplasts at 250*g* for 3–4 min, and resuspend *gently* in 15 mL of STC. Wash the cells again in 15 mL of STC and resuspend in 2 mL of STC.

7. For transformation, mix 100 µL of spheroplasts with the appropriate DNA (up to 1.0 µg in 10 µL or less) and let sit at room temperature for 10 min. Addition of carrier DNA (calf-thymus or herring-sperm) to the sample DNA will protect against nucleases and will increase transformation efficiencies when the concentration of the sample DNA is low.
8. Add 1 mL of PEG solution and let sit at room temperature for 10 min more.
9. Pellet the spheroplasts (250*g*, 3–4 min) and gently resuspend in 150 µL of SOS. Incubate at 30°C for 20–40 min.
10. Mix with TOP agar, kept at 48–50°C, and plate immediately onto a SORB plate lacking uracil. Use 8 mL of TOP for a 90-mm plate and 15 mL for a 132-mm plate.
11. Incubate for 3–4 d at 30°C.

3.5. Characterization of the Library (see *Note 6*)

Once a library has been constructed, it is necessary to characterize it according to the following criteria: percentage of inserts, average insert size, and number of transformants. Following transformation and subsequent growth in single drop-out media, colonies are transferred to the surface of a plate lacking both uracil and tryptophan. Colonies that grow red on this media should contain an artificial chromosome inserted in the *SUP* gene of the vector and flanked by the right and left arms appropriately. However, we have seen red clones that subsequently have been shown to contain no inserts. As an accurate check of percentage of inserts, it is therefore advisable to pick a random selection of clones, prepare agarose blocks as described in the "miniprep" protocol below, and analyze by PFGE. With appropriate size markers, this will also give an estimate of the average insert size. In many cases, the artificial chromosome, when stained using ethidium bromide, will be visible as an additional band on the background of endogenous yeast chromosomes. However, not all artificial chromosomes will be detected by staining, and it is advisable to blot the gel and probe with either pBR322 (which will detect YAC vector sequences) or with total genomic source DNA (or appropriate repeat fragments). Figure 6 gives an example of both a gel stained with ethidium bromide and a gel blotted and probed with pBR322.

3.5.1. "Miniprep" Protocol

1. Pick one colony into 2.5 mL of SORB medium lacking uracil, and grow at 30°C for 36–48 h (OD_{660} should be about 3.0).
2. Pellet at 500*g* (bench centrifuge) for 5 min.
3. Resuspend in 1 mL of 1 *M* sorbitol and transfer to an Eppendorf tube.

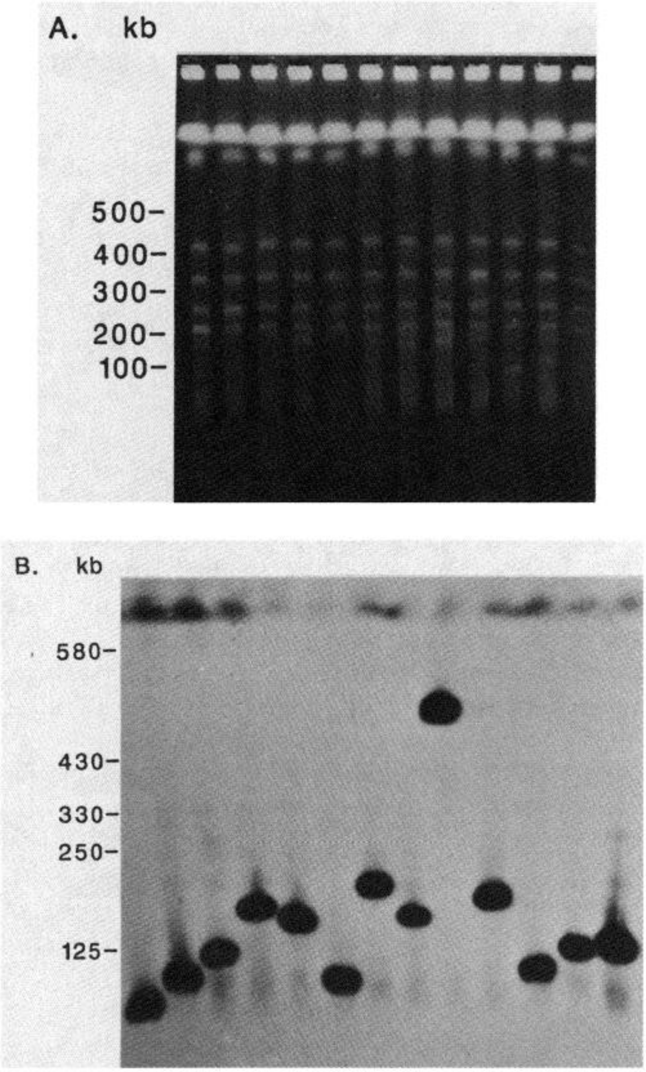


Fig. 6. Analysis of YAC "minipreps" prepared from a random selection of colonies. **(A)** Gel stained with ethidium bromide. **(B)** DNA was transferred from the gel and hybridized with ^{32}P-labeled pBR322, showing the position of the YAC clones. The AB1380 yeast chromosomes are the same in each lane, whereas the YACs vary in size. Gels (0.8% agarose) are run for 22 h, 35-s pulse, 0.5× TBE. Figures A and B show different gels.

4. Pellet and resuspend in 1 mL of 100 m*M* EDTA, pH 8.0.
5. Leave at room temperature for 30 min.
6. Pellet and resuspend in 1*M* sorbitol.
7. Pellet again and resuspend in 90 µL of SCEM. Add 6 µL of 2 mg/mL lyticase in SCE and incubate at 30°C for 1 h.
8. Prepare 1% LMP agarose in 1*M* sorbitol and hold at 55°C.
9. Mix 90 µL of spheroplast suspension with 190 µL of LMP agarose and make two blocks.
10. Incubate overnight at 50°C in 2.5 mL of 20 mg/mL proteinase K; 0.25*M* EDTA, pH 8; and 1% sarkosyl.

    Note: The protocol may be scaled up if more blocks are needed. If blocks are to be used for cloning and mapping, they need to be treated with PMSF prior to restriction-enzyme digestion.

3.6. Screening

There is still much debate over the most efficient way of preparing YAC libraries for screening. This is because no alternative to growing transformed spheroplasts embedded in agar has been widely accepted to date. The first YAC libraries were gridded and stored as individual clones *(7)*. This involves picking each colony from within the agar (using a toothpick or equivalent) and transferring it to the surface of a double drop-out plate in a gridded fashion. Appropriate clones are then transferred to the wells of microtiter plates and subsequently inoculated onto nylon filters in preparation for screening *(see below)* (Fig. 7). The advantages of this method are that it is very simple to accurately retrieve pure single-positive clones and that the library can be stored indefinitely at –80°C, thus making it a permanent resource. In addition, it is relatively simple to make DNA from pools of the clones from each plate and to use PCR to identify the plate that contains the clone of interest *(see below)*. A gridded library is a labor-intensive task, but it is conceivably worthwhile, if it is for an entire genome and can be distributed for screening with many probes.

Methods have now been developed that make it possible to treat YAC libraries much like cosmid libraries, by growing the colonies on the surface of plates and lifting them onto nylon membranes. One such method involves the use of a metal pin-cushion device to pick up the colonies from within the agar and then to inoculate them onto the surface of plates (Lehrach personal communication). More recently, a genomic library has been prepared by plating transformed spheroplasts in alginate, as discussed in Section 3.4 *(26–28)*. Such variations are most suitable when one is looking for a subset of clones in a nonhomologous background or when one constructs a transient library for the purpose of cloning a single or only a few specific YACs, though

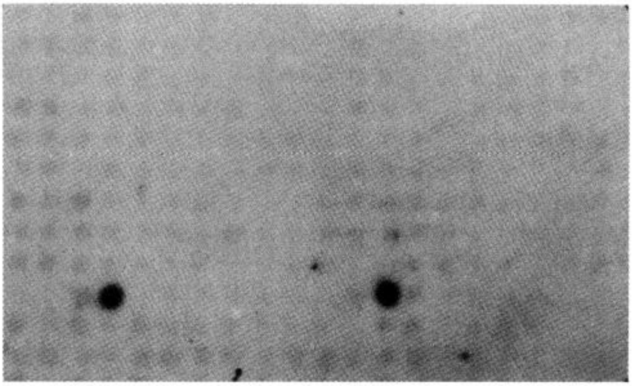


Fig. 7. High-density screening of YAC transformants gridded on Hybond-N nylon filters. In this case, colonies were transferred from microtiter plates, inoculated onto filters in quadruple density, and treated as described in Section 3.6.1. Filters were hybridized by standard methods with ^{32}P-labeled probe, and the positive signals seen were obtained after 3 d of autoradiography. The background grid is useful in identifying the positions of the positive clones.

it has been used very effectively for screening a total human genomic library. A procedure for colony hybridization is included in the following protocol.

3.6.1. Preparation of Colonies for Screening

1. If colonies are to be picked individually for a gridded library, remove them using a toothpick and streak in a gridded array on an SD$^-$ plate (we usually streak 100 clones/90-mm plate).
2. After 2–4 d at 30°C, transfer those clones that have become red to the wells of microtiter plates containing 150 µL of AHC$^-$ and incubate, with shaking, at 30°C.
3. When the cultures have grown (1–2 d), add 50 µL of 80% glycerol to each well using a multichannel pipet. At this stage, it is also possible to make replica plates using a 96-well prong device to inoculate a fresh microtiter plate (containing AHC$^-$) and treating these plates as described above. It is advisable to have several copies of each plate, since the clones will not survive indefinite rounds of freezing and thawing. The plates should be stored frozen at –80°C.
4. To replicate colonies onto filters, remove plates from the freezer and allow them to thaw. Using a 96-well prong device, inoculate a Hybond-N filter and transfer to an AHC$^-$ plate. Incubate at 30°C until the colonies have grown (2–4 d).

5. At this stage, replicas can be made by conventional methods of placing a prehydrated filter over the colonies and applying pressure with a glass plate. The replicas are then regrown on fresh AHC$^-$ plates at 30°C. If desired, the original filter may be stored frozen at –80°C by incubating for 30 min at room temperature on 3MM paper saturated with AHC$^-$ +20% glycerol. Transfer the filter to an empty Petri dish and store frozen.
6. When filters are to be prepared for screening, whether they have been treated as above or by other means, such as use of a pin-cushion device, place the filter in a Petri dish on 3MM paper saturated with SCEM containing zymolyase-100T (50 U/mL). Seal the dish and incubate at 30°C for 1–2 d.
7. Place the filter colony-side-up on 3MM paper saturated with 10% SDS. Incubate at room temperature for 5 min.
8. Transfer to 3MM paper soaked in denaturing solution for 10 min at room temperature, followed by two 5-min incubations on 3MM paper soaked in neutralizing solution.
9. Transfer to a dish of 3× SSC (0.45*M* NaCl, 45 m*M* Tri-sodium citrate, pH 7.0) and wipe off any cell debris with a tissue.
10. Allow to air-dry, and bake in a vacuum oven for 2 h.
11. Subsequent hybridizations are performed as for standard Southern-blotting procedures (*see* Chapter 15).

3.6.2. Preparation of Yeast DNA for PCR Screening

1. Scrape colonies from filters (prepared as above) using the appropriate amount of AHC$^-$. For approx 400 colonies, we scrape into 50 mL of medium.
2. Pellet at 5000 rpm for 5 min and wash the pellet once in SCE.
3. Resuspend in 3 mL of SCEM, 0.25 mg/mL zymolyase-100T, and incubate for 1 h at 30°C.
4. Pellet spheroplasts at 5000 rpm for 5 min, and resuspend in 5 mL of 50 m*M* Tris, pH 7.4; 20 m*M* EDTA; and 1% SDS. Lyse cells by incubation at 65°C for 30 min.
5. Add 1.5 mL of 5*M* potassium acetate and let sit on ice for 60 min.
6. Centrifuge at 10,000 rpm for 10 min.
7. Transfer the supernatant to a fresh tube and precipitate the DNA with 2 vol of ethanol at room temperature. Pellet at room temperature at 10,000 rpm for 10–15 min.
8. Dry the pellet and resuspend in 3 mL of TE, pH 7.4.

9. Centrifuge at 10,000 rpm for 15 min and transfer the supernatant to a fresh tube. Add 0.05 mg/mL RNase and incubate at 37°C for 30 min.
10. Reprecipitate the DNA in 1 vol of isopropanol, spool out, and resuspend in 0.5 mL of TE. If the solution is milky, precipitate it again with isopropanol.

3.7. Analysis of Positive Clones

Once a positive clone has been identified, it should be streaked out for single colonies and then prepared in agarose blocks from overnight cultures of at least four such colonies. These blocks are then used to determine the size of the YAC using PFGE and will also show whether it has been cotransformed with another YAC. When several positive clones have been identified with the same probe, it is important to establish the amount of overlap between them. This can be done by digesting the YAC blocks with one or more frequent cutters (e.g., EcoRI, if it has been used in the cloning; HindIII; or PvuII), performing agarose gel electrophoresis and Southern blotting (Fig. 8). These can then be hybridized with moderate and highly repetitive elements, or with total human genomic DNA, to give a good idea of the amount of overlap between clones.

Restriction maps of the YAC clones are very useful and can be generated in a variety of ways. The YACs may be digested in blocks with rare cutters and the fragments first resolved by PFGE or FIGE and then Southern blotted (Fig. 9). They are then hybridized sequentially with pBR322 sequences specific for each of the YAC arms. These are obtained by digesting pBR322 with PvuII and BamHI and cutting 1.7-kbp (specific for the right arm) and 2.6-kbp (specific for the left arm) fragments from a LMP gel *(2)*. Restriction maps of frequent cutters can be generated by indirect end-label mapping *(30)*.

It is often necessary to obtain the end fragments of YACs—for example, when this fragment is to be used for directional walking or for extending a PF map. In the pYAC series, it is possible to retrieve the end on the left-arm side of the YAC by digesting the DNA with XhoI or NdeI, diluting the DNA, ligating, and transforming into *E. coli* using the ampicillin-resistant marker present in the vector arm *(2)*. The disadvantages are that this will work only in cases in which suitably placed XhoI or NdeI sites are present and will retrieve only one end of the YAC. Vectors that make it possible to retrieve both end fragments have been developed *(26)*. An alternative end-cloning technique is to use inverted PCR with primers that are specific for sequences flanking the insert *(9)*.

When YACs are too large to be used for some types of analyses, they can easily be subcloned into plasmid, cosmid, or phage vectors. Melt 3–4 blocks and treat them with agarase as described previously, concentrate to a vol of

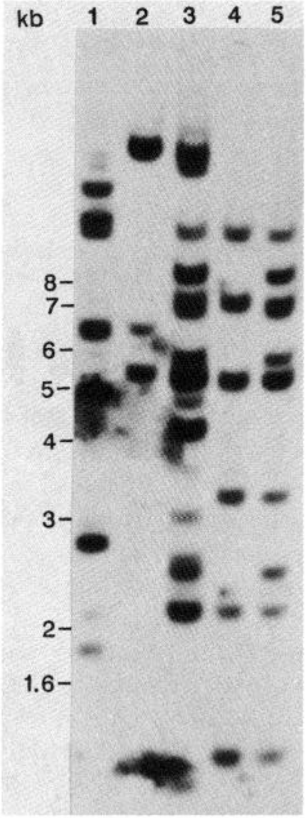


Fig. 8. Autoradiograph of YAC DNA digested to completion with EcoRI and hybridized to radiolabeled total human genomic DNA, illustrating the amount of overlap between YAC clones obtained with the probe G2 in lanes 1 and 2; and with D9 in lanes 3, 4, and 5. The following samples are shown: G2-2 (lane 1), G2-1 (lane 2), D9-3 (lane 3), D9-2 (lane 4), and D9-1 (lane 5). The overlap between the D9 YACs can clearly be seen.

300 µL and dialyze in a collodion bag, do test digestions with MboI (0.001–1 U for 200 ng of DNA), select suitable conditions for a scaleup experiment, and ligate into the appropriate vector. The resultant clones will originate either from the YAC or from the endogenous yeast chromosomes. Clones resulting from the YAC can be identified from the background of yeast clones by screening with a repeat sequence probe. Alternatively, if the artificial chromosome runs clear of the endogenous yeast chromosomes on a PFG, the band can be cut from the gel and cloned directly. Large YACs may also be

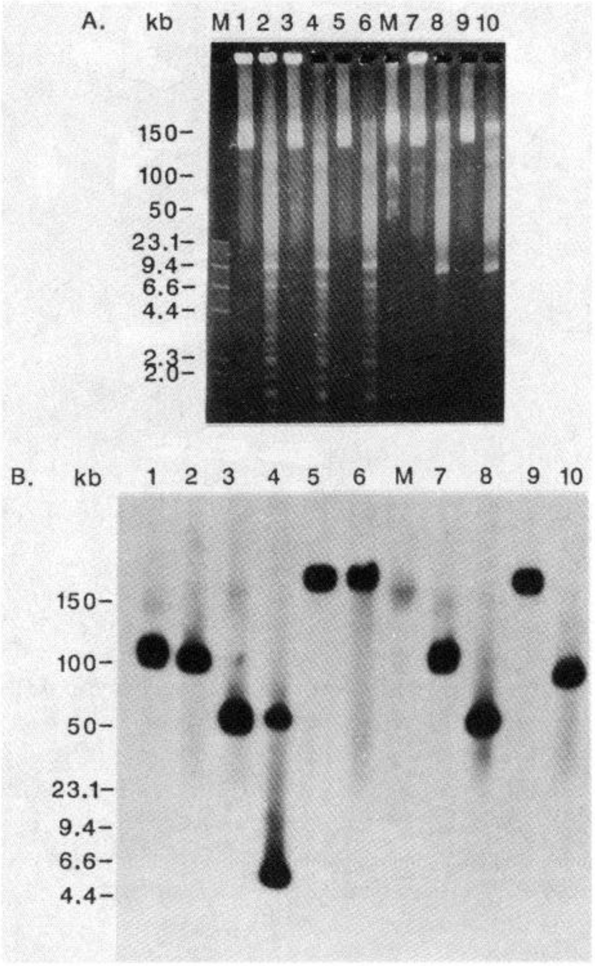


Fig. 9. FIGE gel stained with ethidium bromide (A) prepared with the following YAC blocks: D9-1 (lanes 1, 2, 7, and 8); D9-2 (lanes 3 and 4); and D9-3 (lanes 5, 6, 9, and 10). Undigested DNA is present in lanes 1, 3, 5, 7, and 9; DNA digested with SacII in lanes 2, 4, and 6; and BssHII in lanes 8 and 10. The marker on the left is λ DNA digested with HindIII and that in the center, a λ concatamer. A Southern transfer of the gel was hybridized with fragments of pBR322 digested with PvuII and BamHI. The autoradiograph was hybridized (B) with the small fragment (right-arm specific) and (C) (*see* following page) with the large fragment (left-arm specific). D9-1 and D9-3 both contain a BssHII site, but no SacII site. D9-2 has a SacII site that is partly digested (lane 4).

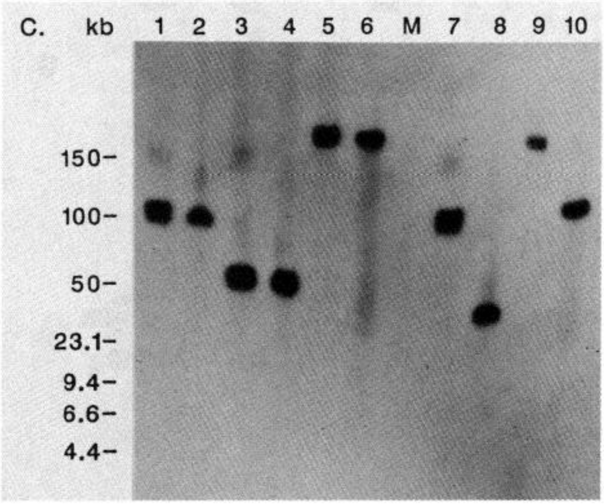


Fig. 9(C).

fragmented by producing deletion derivatives by targeted homologous recombination using a human Alu family repeat sequence *(21,31)*.

The aim of many YAC cloning experiments will be to identify coding sequences, and this can be achieved by cloning CpG-hypomethylated islands, looking for cross-species hybridization or expression in appropriate tissues using Northern blots, and screening cDNA libraries. CpG-hypomethylated islands can be cloned by using a vector with suitable cloning sites. The entire YAC may be excised from LMP agarose, preassociated with genomic DNA, and used directly to probe Northern blots. This procedure should even be suitable for screening cDNA libraries provided sufficient preassociation with vector sequences is carried out.

4. Notes

1. Care of yeast strains. Yeast strains are prone to alter their phenotype when they are stored on plates for long periods of time. It is therefore advisable to freeze the strains at –80°C in 20% glycerol and streak a fresh plate from the glycerol every 4 wk. For a general introduction to yeast experimentation, consult ref. *32* or *33*. When specific YAC clones are identified, they should also be stored frozen.
2. Genotype of AB1380: MATa$^+$ ura3 trp1 ade2-1 can1-100 lys2-1 his5.
3. When manipulating DNA of very high mol wt, be sure to minimize the number of pipeting steps and to use disposable pipet tips from which the tip has been cut off using a sterile blade in order to enlarge the bore.

4. Check the stability of the vector prior to freezing stocks and preparing bulk vector DNA. The telomeric repeat sequences may rearrange or delete. The integrity of the vector can be tested by digestion with HindIII, in which case, bands of 3.5, 3.0, and 1.9 kbp, and a doublet at 1.4 kbp are expected. If the telomere has been deleted, the doublet will be resolved and five bands will be seen.
5. If the agarose does not dissolve, increase the salt concentration in order to melt the block, but be sure to decrease the concentration prior to transformation, since it may decrease the transformation frequency.
6. Complexity of the library. The number of clones needed if three genome equivalents are to be screened with a 90% chance of finding a particular single-copy sequence in a genome of 3×10^9 bp.

Average insert size, kbp	Number of transformants
100	90,000
200	45,000
300	30,000
400	22,500

7. It is worth noting that DNA prepared in liquid will run as a smear, often extending down to 300 kbp even before digestion. This can make detection of appropriate partial digestion levels difficult.
8. Some groups suggest phenol/chloroform extraction of DNA before ligation and again after ligation, prior to long-term storage. If DNA is to be extracted at any stage, it is essential to use high-quality phenol and to be extremely gentle. The suggested method involves adding an equal vol of freshly equilibrated phenol slowly down the side of the tube containing the DNA sample. The sample tube is left on its side for 15 min, and then placed for 10 min in an upright position. The aqueous phase and the interface are removed after a 5-s spin in a microfuge. The procedure is repeated with chloroform.

References

1. Murray, A W and Szostak, J W. (1983) Construction of artificial chromosomes in yeast *Nature* **305,** 189–193
2. Burke, D. T , Carle, G F., and Olson, M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236,** 806–812.
3. Heard, E., Davies, B , Feo, S., and Fried, M. (1989) An improved method for the screening of YAC libraries *Nucleic Acids Res.* **17,** 857.
4. Garza, D , Ajioka, J. W., Burke, D T , and Hartl, D L. (1989) Mapping the *Drosophila* genome with yeast artificial chromosomes. *Science* **246,** 641–646.

5. Coulson, A., Waterston, R , Kiff, J , Sulston, J., and Kohara, Y. (1988) Genome linking with yeast artificial chromosomes. *Nature* **335**, 184–186.
6. Little, R. D., Porta, G., Carle, G. F., Schlessinger, D., and D'Urso, M. (1989) Yeast artificial chromosomes with 200- to 800-kilobase inserts of human DNA containing HLA, V_{κ}, 5S, and Xq24–Xq28 sequences *Proc. Natl. Acad. Sci USA* **86**, 1598–1602.
7. Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J , Schlessinger, D., and Olson, M. V. (1989) Isolation of single-copy human genes from a library of yeast artificial-chromosome clones. *Science* **244**, 1348–1351.
8. Guzman, P. and Ecker, J. R. (1988) Development of large DNA methods for plants: Molecular cloning of large segments of Arabidopsis and carrot DNA onto yeast *Nucleic Acids Res.* **16**, 11091–11105.
9. Silverman, G. A., Ye, R D , Pollock, K. M., Sadler, J E , and Korsmeyer, S J (1989) Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21 3. *Proc. Natl. Acad. Sci. USA* **86**, 7485–7489.
10. Burmeister, M. and Lehrach, H. (1986) Long range restriction map around the Duchenne muscular dystrophy gene. *Nature* **324**, 582–585.
11. Gitschier, J , Wood, W. I , Goralka, T. M., Wion, K. L , Chen, E Y, Eaton, D. H , Vehar, G. A., Capon, D. J , and Lawn, R M (1984) Characterisation of the human factor VIII gene. *Nature* **312**, 326–330
12. Rommens, J. M., Iannuzzi, M., Kerem, B.-S., Drumm, M. L., Melmer, G., Dean, M , Rozmahel, R , Cole, J L , Kennedy, D , Hidaka, N , Zsiga, M , Buchwald, M., Riordan, J R , Tsui, L.-C., and Collins, F. S. (1989) Identification of the cystic fibrosis gene Chromosome walking and jumping. *Science* **245**, 1059–1065.
13. Bennet, J. and Gearhart, J. D. (1989) Transomic mice—microinjection of large cloned DNA sequences. Abstract from Gordon conference, June, 1989, New Hampshire.
14. Marchuk, D and Collins, F. S. (1988) pYAC-RC, a yeast artificial chromosome vector for cloning DNA cut with infrequently cutting restriction endonucleases *Nucleic Acids Res* **16**, 7743.
15. Cooke, H. and Cross, S. (1988) pYAC-4 Neo, a yeast artificial chromosome vector which codes for G418 resistance in mammalian cells. *Nucleic Acids Res.* **16**, 11817.
16. Cross, S. H., Allshire, R. C., McKay, S. J., McGill, N. I., and Cooke, H. J (1989) Cloning of human telomeres by complementation in yeast. *Nature* **338**, 771–774.
17. Brown, W. R. A (1989) Molecular cloning of human telomeres in yeast. *Nature* **338**, 774–776.
18. Riethman, H C , Moyzis, R. K., Meyne, J , Burke, D T., and Olson, M. V. (1989) Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. *Proc Natl. Acad Sci USA* **86**, 6240–6244
19. Hahnenberger, K. H., Baum, M P , Polizzi C M , Carbon, J , and Clarke, L. (1989) Construction of functional artificial minichromosomes in the fission yeast *Schizosaccharomyces pombe* *Proc. Natl. Acad Sci. USA* **86**, 577–581.
20. McCormick, M K., Shero, J. H., Connelly, C. J., Antonarakis, S. E., and Hieter, P A (1990) Methods for cloning large DNA segments as artificial chromosomes in *S cerevisiae. Technique* **2**, 65–71
21. Pavan, W J , Hieter, P., and Reeves, R. N. (1989) Generation of deletion derivatives by targetted transformation of human-derived yeast artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **87**, 1300–1304

22. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
23. Anand, R., Villasante, A., and Tyler-Smith, C. (1989) Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed-field gel electrophoresis. *Nucleic Acids Res.* **17,** 3425–3433.
24. Burke, D. T. and Olson, M. V. Preparation of clone libraries in yeast artificial chromosome vectors. *Methods Enzymol.,* in press.
25. Burgers, P. M. J. and Percival, J. (1987) Transformation of yeast spheroplasts without cell fusion. *Anal. Biochem.* **163,** 391–397.
26. Traver, C. N., Klapholz, S., Hyman, R. W., and Davis, R. W. (1989) Rapid screening of a human genomic library in yeast artificial chromosomes for single-copy sequences *Proc. Natl. Acad. Sci. USA,* **86,** 5898–5902.
27 Lambie, E J and Roeder, G. S. (1986) Repression of meiotic crossing over by a centromere (CEN3) in *Saccharomyces cerevisiae. Genetics* **144,** 769–789.
28. Lai, E. and Cantrell, C. (1989) Rapid colony screening of YAC libraries by using alginate as matrix support *Nucleic Acids Res.* **17,** 8008.
29 Ito, H , Fukuda, Y., Murata, K., and Kimura, A. (1983) Transformation of intact yeast cells treated with alkali cations. *J Bacteriol.* **153,** 163–168.
30. Church, G. M. and Gilbert, W. (1984) Genomic sequencing *Proc. Natl. Acad. Sci. USA* **81,** 1991–1995.
31. Vollrath, D., Davis, R. W , Connelly, C , and Hieter, P. (1988) Physical mapping of large DNA by chromosome fragmentation. *Proc. Natl. Acad Sci. USA* **85,** 6027–6031
32. Campbell, I. and Duffus, J H., eds. (1988) *Yeast — A Practical Approach.* IRL, Oxford, UK
33. Sherman, F , Fink, G. R., and Hicks, J B. (1986) *Laboratory Course Manual for Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY)
34. Cooke, H. (1987) Cloning in yeast. An appropriate scale for mammalian genomes. *Trends Genet.* **3,** 173,174
35. Hieter, P., Mann, C., Snyder, M , and Davis, R. (1985) Mitotic stability of yeast chromosomes: A colony colour assay that measures nondisjunction and chromosome loss. *Cell* **40,** 381–392.

CHAPTER 20

Gene Targeting for Somatic Cell Manipulation

Julia R. Dorin and David J. Porteous

1. Introduction

Recent technological advances, not least in somatic cell genetics, have accelerated progress in human genome mapping and reverse genetics. In particular, chromosome mediated gene transfer (CMGT) *(1)* and irradiation and fusion gene transfer (IFGT) *(2)* have been exploited to great effect for enrichment cloning of predetermined regions of the human genome. Both methods allow subchromosomal fragments of human DNA (generally 1 to 50 Mbp in size) to be isolated onto a rodent cell background as a resource for DNA cloning and fine structure mapping.

The success of these methods relies on a selectable marker mapping close to or within the chromosomal region of interest. A rather limited number of biochemical selection systems can be used to isolate specific regions of the genome, e.g., hypoxanthine phosphoribosyl transferase (HPRT) for Xqter and thymidine kinase (TK) for 17q. The oncogenes provide a further, novel, and widespread class of selectable marker to isolate loci involved in human cancer and linked chromosomal regions *(3)*. In addition, antibodies directed against cell surface antigens *(4)* can be used to select for cells containing the corresponding chromosomal fragments.

However, despite the possibilities described above, no endogenous selectable marker is known for a substantial portion of the human genome. We describe here a solution to this limitation by targeting an exogenous domi-

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

nant biochemically selectable marker precisely to a predetermined locus *(5)*. The selection is based on the antibiotic resistance gene neomycin phosphotransferase *(neo)*. This gene has been used in a wide variety of cultured mammalian cells to confer resistance to the antibiotic Geneticin (G418). Normally, the *neo* gene is placed under an ubiquitous eukaryotic promoter (commonly from SV40 *[6]*) so that expression is not dependent on the integration site. If a length of cloned DNA is included in the vector, homologous recombination with the endogenous target sequence will direct the *neo* gene to the desired chromosomal region. Unfortunately, in mammalian cells, this process is extremely inefficient and the ratio of nonhomologous recombination to homologous recombination events is about 1000:1 *(7,8)*. Certain experimental refinements can improve this ratio, to, at best, 100:1. These include using a targeting vector with the greatest amount of homologous sequence possible; generating DNA ends within the homologous region; and using target cells that are in early S phase *(9)*. Our method stringently selects for correctly targeted events by fusing the *neo* gene without any promoter sequences, in frame, to the coding sequences of a gene within the chromosomal target. Thus, *neo* expression depends on endogenous promoter function and read-through transcription. The essential requirements are, thus, a cloned and characterized gene within the region of interest and a transfectable, expressing cell line.

The first stage of our approach is to design a suitable targeting vector with a promoterless *neo* cassette coupled to a region of homology from the target gene. Next, the vector is introduced into the relevant cell line. Resultant G418 resistant clones are analyzed by the polymerase chain reaction (PCR) and Southern blotting to identify and confirm the correctly targeted event. The targeted *neo* can then be used in any variety of gene transfer experiments to generate subchromosomal somatic cell hybrids retaining only the region of ultimate interest. These may then be subjected to exhaustive functional and molecular analysis to achieve the ultimate objective of the experiment. It should be mentioned that having isolated candidate genes from a particular area of the genome, gene targeting can be used again to provide a unique insight into biological function (*see* Note 1). We will describe the isolation of regions of human chromosome 7 as an example.

We chose the SV40 gene array in the human–mouse hybrid C121 as our target locus. The only human chromosome in this hybrid is chromosome 7, and the SV40 has independent integration sites at 7q11 (unpublished observation) and 7q31 *(10)*. The large *T antigen* gene of SV40 is highly expressed in C121 and so, this gene was used as the gene target. Using our targeting vector pTAGNEO in the following protocol, we obtained three correctly targeted clones out of 11 G418 resistant colonies isolated (*see* Note 2). This frequency,

together with those obtained by other workers *(11–13),* indicates that using an enhancer- or promoter-deficient gene provides a stringent selection for homologous recombination events.

2. Materials

1. Targeting vector at 1–2 mg/mL in TE (10 m*M* Tris-HCl, pH 7.4, 0.1 m*M* EDTA).
2. Sterile tissue-culture flasks and plates.
3. The electroporation machine used here was made in our Medical Research Council (MRC) workshop, but the Bio-Rad gene pulser and capacitance extender (or other commercial alternative) are a suitable substitute.
4. Autoclaved silicone grease and cloning rings (*see* Note 3).
5. Tissue-culture medium: Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum.
6. Selective tissue-culture medium: Tissue culture medium plus 800 µg/mL Geneticin (G418) (*see* Note 4).
7. Freezing mix: 10% Dimethyl Sulfoxide in fetal calf serum.
8. Phosphate buffered saline (PBS): 170 m*M* NaCl, 3 m*M* KCl, 10 m*M* $Na_2HPO_4.12H_2O$, 1.8 m*M* H_2PO_4, pH 7.2.
9. Trypsin/versene: 0.25% porcine trypsin, 1 m*M* EDTA, 120 m*M* NaCl, 2 m*M* $Na_2HPO_4.12H_2O$, 1.7 m*M* KH_2PO_4, 5 m*M* KCl, 5 m*M* D-glucose, 24 m*M* Tris, pH 7.6. Filter through a 0.22 µ*M* filter and store at –20°C.
10. Chromosome buffer: 15 m*M* Tris-HCl, 3 m*M* $CaCl_2$, pH 7.0.
11. 2× HeBS: 50 m*M* HEPES, 280 m*M* NaCl, 1.50 m*M* Na_2HPO_4, pH 7.15.
12. 2× $CaCl_2$: 250 m*M* $CaCl_2$.
13. Colcemid: 0.1 mg/mL in double-distilled H_2O (1000× solution).
14. 0.075*M* KCl made in deionized H_2O.
15. 0.1% Digitonin (*see* Note 5) in chromosome buffer.
16. 50% sucrose in double distilled water.
17. 5% Sucrose in chromosome buffer.
18. 15% Glycerol in 1× HeBS.

3. Methods

3.1. Vector Design

In this example, the promoterless *neo* gene containing the HSV-tk poly(A) addition site is fused to the majority of the *T antigen* gene (*see* Note 6), so that reading from the *T antigen* codons runs in-frame into the *neo* gene coding

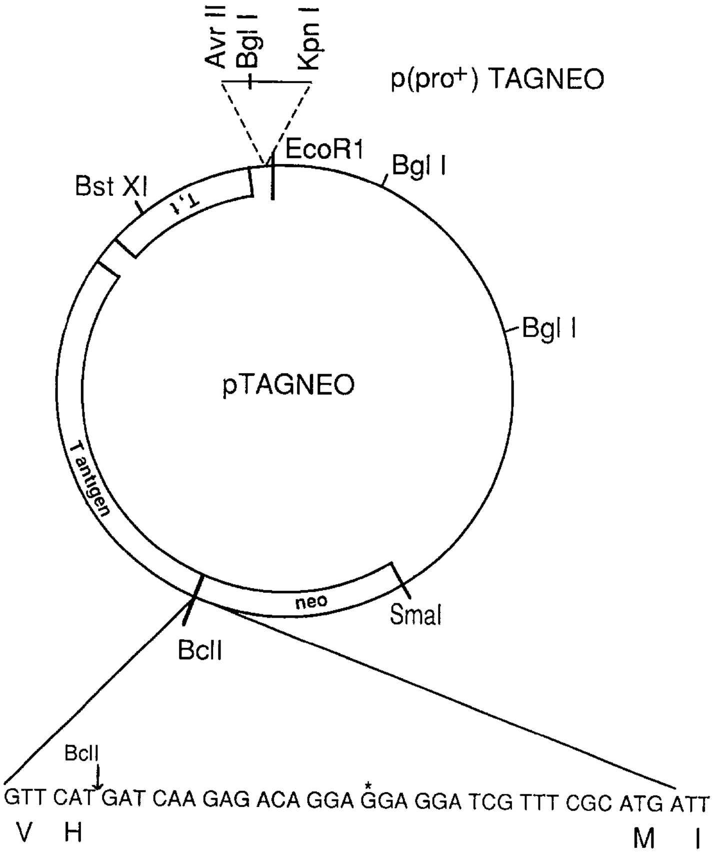


Fig. 1. The SV40 targeting vector with (p(pro$^+$) TAGNEO) and without (pTAGNEO) promoter sequences. The construct is as described in the text.

region. The sequence illustrated in Fig. 1 shows the last two codons of the *T antigen* portion, the cloning linker sequence with a stop codon removed (asterisked), and the methionine at the beginning of *neo* (*see* Note 7). The unique Bst XI site in the region of homology is used to linearize the vector prior to introduction into the cell (*see* Note 8). It is important, of course, to determine whether or not the *neo* fusion protein will function and confer G418 resistance on cells expressing the protein. To this end, we constructed p(pro$^+$)TAGNEO, which is identical to pTAGNEO except that it includes the SV40 early promoter and enhancer sequences upstream of the *T antigen* gene, and should therefore confer G418 resistance on cells at a high frequency (irrespective of integration site) if the fusion protein is functional.

3.2. Electroporation

Generally speaking, targeting vectors are introduced into cells either by microinjection or electroporation. For reasons of experimental ease, electroporation is definitely the method of choice where there is a selection for the vector (*see* Note 9). A good but not absolute guide of effective electropermeabilization is that the pulse should result in 40–80% cell death (*see* Note 10).

1. Grow the C121 cells in four 175-cm tissue-culture flasks in tissue-culture medium and harvest when the cells are 50–80% confluent. To harvest the cells, wash out the flasks with PBS and then add 5 mL of trypsin/versene for 5 min at 37°C.
2. Shake the cells off and remove the trypsin/versene cell mixture into a conical-bottomed centrifuge tube. Wash the cells in culture medium and then in PBS by pelleting at 500*g* and resuspending in 10 mL of liquid. Count the cells with a hemocytometer slide and resuspend in PBS on ice at a concentration of 10^7 cells/mL. Prepare five identical 0.5 mL cell aliquots.
3. To the five tubes, add 10–20 µL of a solution containing (i) no DNA, (ii) Bst XI linearized pTAGNEO, (iii) Bst r XI linearized p(pro$^+$) TAGNEO, (iv) intact pTAGNEO, or (v) intact p(pro)+TAGNEO (*see* Notes 11 and 12).
4. Place each tube's contents into an electroporation cuvet and expose to an electric pulse of 3000 V/cm at room temperature (*see* Note 13).
5. Allow cells to recover for 10 min on ice and then plate out each experiment into a 175 cm^2 tissue-culture flask in tissue-culture medium.
6. After 24 h, put the cells onto selective tissue-culture medium. After approx 10–14 d growth (changing the medium every third day), G418 resistant clones should be evident by eye when the flask is held up to the light. Clone numbers should be recorded and clones generated by BST XI cut pTAGNEO screened for homologous recombination (*see* Note 14).
7. Wash the flasks out with PBS and isolate clones by cutting the top off the flask with a soldering iron and using a cloning ring. Place the ring into a dish of autoclaved silicone grease and then put the ring over the clone so that the grease causes a seal against the flask. Fill the ring with a couple of drops of trypsin/versene and wait 5 min. Remove the trypsin/versene with a Pasteur pipet and add to a 2-mL tissue-culture well filled with selective tissue-culture medium.
8. Wash out the ring with the medium and add to the well. Put one-half of the medium into a universal and spin down the cells at 500*g* for 5 min.

These cells are for PCR screening for homologous recombinants (*see* Note 15 and Chapter 1 for method). Allow the remaining cells to grow, and harvest the clones that are identified as positive homologous recombinants by PCR when the 2-mL well is confluent. Spin down one-half of the cells and resuspend in 1-mL ice-cold freezing mix and put at –70°C in a freezing vial. After 24 h, transfer to liquid nitrogen for long-term storage. Alternatively, viable cells may be maintained at –70°C for several weeks while the experiment is analyzed. Spin down the other cells and resuspend in 10 mL of tissue-culture medium and put into a 25-cm^2 tissue culture flask. When this flask is confluent, harvest 90% of the cells and prepare DNA to validate the PCR result.

3.3. Chromosome Mediated Gene Transfer (CMGT)

Once a clone has been identified that has *neo* integrated into the chromosomal region of choice, it is possible to isolate that subchromosomal fragment onto a background of another species using CMGT or IFGT and G418 selection. We have used pTAGNEO to target into two sites of human chromosome 7 and then enriched for different subchromosomal regions of that chromosome on a C127 mouse cell background by using CMGT (*see* refs. *3, 14,* and *15* for further details), as follows.

1. Arrest the donor cells in mitosis by the addition of Colcemid (Caution: Colcemid is a mutagen) (0.1 µg/mL for 18 h) to the culture medium.
2. Shake off mitotic cells, chill on ice and pellet by centrifugation (200*g* for 7 min).
3. Resuspend to 10^6 cells/mL in fresh 0.075*M* KCl (in deionized water) and leave to swell (10–20 min at 37°C or 30–40 min at room temperature). Chill on ice and spin down gently (150*g* for 7 min) at 4°C. Aspirate the supernatant and resuspend to 5 × 10^6 cells/mL in 0.1% digitonin (Caution: Digitonin is toxic) in chromosome buffer. Keep at 4°C until Step 7.
4. Gently draw the cell suspension three or four times through a 21-gage needle. Check for chromosome release and integrity by light microscopy. Aim for a final concentration of between 5 × 10^7 and 2 × 10^8 chromosomes (equivalent to 6–24 µg DNA) in 0.8 mL coprecipitate/plate.
5. Remove intact nuclei, whole cells, and debris by centrifugation (7 min at 100*g*). Dilute out the digitonin with chromosome buffer. Add sucrose to 5% (to prevent clumping and ease resuspension of chromosomes) and pellet the chromosomes by centrifugation at 1300*g* for 20 min.

6. Aspirate off all the supernatant and gently resuspend the chromosomes in 5% sucrose in chromosome buffer. Repellet the chromosomes by centrifugation at 1300*g* for 20 min.
7. Aspirate off all the supernatant and resuspend the chromosomes gently and evenly in 2× HeBS at room temperature in a clear plastic centrifuge tube (between 10^8 and 4×10^8 chromosomes in 0.4 mL 2× HeBS). Add an equal vol of 2× $CaCl_2$, dropwise, gently bubbling air through to mix. A bluish and lightly flocculent coprecipitate of chromosomes and calcium phosphate will form.
8. Aspirate the medium from the recipient cells (5×10^5 – 10^6/100-mm dish, seeded 18 h previously) and add the freshly formed coprecipitate (0.8 mL/plate).
9. Leave for 20 min at room temperature, agitating occasionally.
10. Add 10 vol of tissue-culture medium (*see* Note 16) and incubate at 37°C for 6–8 h.
11. Aspirate the medium and coprecipitate and shock the cells with glycerol (15% in HeBS, 2–3 min at 37°C). Aspirate off the glycerol, wash once with tissue-culture medium, and refeed with the same.
12. After 24 h, supplement the medium with 800 µg/mL G418 and, after approx 14–21 d, pick clones as before.

    The human DNA content can be assessed on a Southern blot using a human-specific dispersed repeat (L1 fingerprinting—ref. *15*) and the extent and fidelity of cotransfer of DNA using species-specific synthetic DNA probes. *In situ* hybridization can also be used to visualize the transgenome directly *(10)*.

4. Notes

1. Gene targeting loci has intriguing possibilities in terms of disruption of specific loci leading to the examination of gene function at the cellular or whole animal level. Gene targeting in mouse embryonic stem cells (ES) is a powerful genetic tool as, after manipulation, these cells can be injected into blastocysts from which somatic, and more rarely, germ-line chimeras arise *(16)*. These mice can be used to examine gene function, or to provide models for human genetic disease and possibly gene therapy. Targeting into a gene in ES cells is technically much more demanding than other cell types, as the cells must be maintained in an undifferentiated state if they are to contribute effectively to the somatic tissue, and eventually to the germ line. Targeting vectors designed in the

way we describe will only be appropriate for targeting genes expressed in the ES cells. Microinjection and PCR screening *(17)* is one solution, but Mansour et al., 1988 *(18)* have described a very elegant positive/negative selection (PNS) that should be applicable to any gene.

2. The cell line C121 used in this instance contained several copies of human chromosome 7 with an SV40 block consisting of four intact viral genomes in a tandem array. It should be noted that copy number is not a limiting factor in homologous recombination *(19)* and so, although T antigen itself may have enhanced the homologous recombination frequency, the increased copy number is not a relevant factor.
3. Cloning rings can be easily made by cutting off the top 1 cm of a blue Gilson tip.
4. The concentration of G418 required to kill a particular cell line should be determined for each batch. Some cell lines are much more sensitive than others, and for purely economic reasons, it is a good idea to use as low a concentration as possible. As a guide, expect the optimal killing concentration to fall in the range 150 µg/mL and 1 mg/mL.
5. Digitonin is subject to batch variability, and we recommend Fluka.
6. The vector was engineered by standard restriction endonuclease digestion and ligation.
7. The promoterless *neo* can be ligated to any gene in principle, but care should be taken that the reading frames are compatible, and if manipulation of restriction sites is necessary to achieve this, then we advise confirmation by DNA sequencing of the junction in the final construct. We had to use site-directed mutagenesis to remove the stop codon lying between the *T antigen* and *neo* coding sequences. This promoterless *neo* is in a Bcl I-Hind III cassette and available from us on request.
8. The majority of the SV40 *T antigen* has been included in the vector and the vector has been linearized within the region of homology in order to optimize the frequency of homologous recombination events.
9. The brief exposure of the cells to an electric pulse temporarily permeabilizes the cell membrane and allows efficient transfection. However, the conditions must be optimized for each cell type. The appropriate field strength and pulse length can only be determined experimentally or gleaned from the literature. Pulse length is the product of the capacitance and resistance. The resistance is dependent on the electroporation medium and volume. A medium based on physiological saline is best, e.g., PBS or culture medium. The voltage-dependence curve can be rather narrow, and 100 V/cm adjustments will be necessary to find the peak.

10. It is important to wait at least 3 h before using a vital dye to assess cell survival, as live cells will remain permeable for sometime after electroporation.
11. The restriction endonuclease-digested DNA should be extracted twice with phenol-chloroform and precipitated from the aqueous phase with 1/10th vol 3*M* NaOAc and 2.5 vol ethanol. DNA should be resuspended at a concentration of 1 mg/mL in TE.
12. Intact p(pro)+TAGNEO was extremely efficient at conferring G418 resistance onto C121 cells (3.6×10^3 colonies/10^6 cells). The linearized pTAGNEO (6 colonies/10^6 cells) gave 100-fold fewer G418 resistant colonies than the linearized p(pro)+TAGNEO (600 colonies/10^6 cells), reflecting the selection for an active promoter and translation of a functional protein.
13. We used a homemade electroporation apparatus that supplied a 3000 V/cm pulse, by 4 µF capacitance direct discharge through 22-ohm resistance in series.
14. G418 resistant clones can be screened either by Southern blot analysis identifying a novel restriction site within the target gene, or by PCR. The advantage of PCR is that the cells can be taken at the time of clone isolation and nonhomologous recombinants can be discarded before the cells are even frozen down.
15. One DNA primer is chosen to hybridize uniquely to sequence provided by the *neo* gene, and the other oligonucleotide to sequences on the opposite strand in the endogenous target gene and not present in the vector. Only if homologous recombination has occurred will the two primer sites be juxtaposed, and the intervening DNA will be exponentially amplified. PCR-positive clones should always be checked by Southern blotting to avoid wasted effort analyzing false positives.
16. DMEM should be used at this step and not RPMI 1640 or similar high phosphate media.

Acknowledgments

Julia Dorin is supported by the Cystic Fibrosis Trust (UK).

References

1. McBride, W. O. and Ozer, H. L. (1973) Transfer of genetic information by purified metaphase chromosomes. *Proc. Natl. Acad. Sci. USA* **70,** 1258–1262.
2. Goss, S. J. and Harris, H (1975) New method for mapping genes in human chromosomes. *Nature* **255,** 680
3. Porteous, D. J. (1987) Chromosome mediated gene transfer: a functional assay for complex loci and an aid to human genome mapping. *Trends Genet.* **3,** 177.

4. Tunnacliffe, A., Jones, C., and Goodfellow, P. N. (1983) Somatic cell genetics, immunogenetics and gene mapping. *Immunol. Today* 4, 230.
5. Dorin, J. R., Inglis, J. D., and Porteous, D. J. (1989) Selection for precise chromosomal targeting of a dominant marker by homologous recombination *Science* **243**, 1357–1360.
6. Southern, P. J. and Berg, P. (1982) Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *J. Mol. Appl. Genet.* 1, 327–341.
7. Smithies, O., Gregg, R., Boggs, S., Koralewski, M., and Kucherlapati, R. (1985) Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230.
8. Thomas, K. and Capecchi, M. (1987) Site directed mutagenesis by gene targeting in mouse-embryo derived stem cells. *Cell* **51**, 503.
9. Capecchi, M. (1989) The new mouse genetics: Altering the genome by gene targeting. *Trends Genet.* **5**, 70.
10. Mitchell, A., Ambros, P., Gosden, J., Morten, J., and Porteous, D (1986) Gene mapping and physical arrangements of human chromatin in transformed hybrid cells. Fluorescent and autoradiographic *in situ* hybridisation compared *Som. Cell Mol. Genet.* **12**, 313.
11. Jasin, M and Berg, P. (1988) Homologous integration in mammalian cells without target gene selection. *Genes Dev.* **2**, 1353.
12. Sedivy, J. and Sharp, P. (1989) Positive genetic selection for gene disruption in mammalian cells by homologous recombination. *Proc. Natl. Acad. Sci. USA* **86**, 227.
13. Jasin, M., Elledge, S., Davis, R., and Berg, P. (1990) Gene targeting at the human CD4 locus by epitope addition. *Genes Dev.* **4**, 157.
14. Morten, J., Hirst, M., and Porteous, D. (1987) The *c*-Harvey-*ras*-1 oncogene in chromosome mediated gene transfer. *Anticancer Res.* **7**, 573.
15. Porteous, D., Morten, J., Cranston, G., Fletcher, J., Mitchell, A , van Heyningen, V., Fantes, J., Boyd, P., and Hastie, N. (1986) Molecular and physical arrangements of human DNA in HRAS-1 selected chromosome mediated transfectants. *Mol. Cell Biol.* **6**, 2223.
16. Thompson, S., Clarke, A., Pow, A., Hooper, M., and Melton, D. (1989) Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells *Cell* **56**, 313.
17 Zimmer, A. and Gruss, P. (1989) Production of chimaeric mice containing embryonic stem cells carrying a homeobox Hox1.1 allele mutated by homologous recombination *Nature* **338**, 150
18. Mansour, S., Thomas, K., and Capecchi, M. (1988) Disruption of the proto-oncogene int-2 in mouse embryo derived stem cells: A general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348.
19 Zheng, H. and Wilson, J. (1990) Gene targeting in normal and amplified cell lines *Nature* **344**, 170.

CHAPTER 21

In Situ Hybridization of Chromosomes

Kong H. Choo, Ruth M. Brown, and Elizabeth Earle

1. Introduction

This technique allows specific DNA sequences to be directly localized on chromosomes. The probe DNA is labeled by one of several methods that are described in this chapter. The labeled DNA is then applied onto chromosomes fixed on slides under conditions that allow specific hybridization of the probe to its complementary sequence on a chromosome. Following hybridization, unbound probe is removed by washing and the distribution of the probe visualized by autoradiography, immunofluorescence, or enzyme staining.

One of the principal applications of this technique is to map DNA sequences to a specific site on a chromosome. More recently, the technique is increasingly being used in conjunction with standard cytogenetic methods to detect chromosomal aberrations. Some examples of these different applications are described under the appropriate headings below.

There are many variations to the basic technique of *in situ* hybridization, depending on specific purpose, availability of tissue or cell type, or the choice of radioactive vs nonradioactive labeling methods. For the general aim of mapping a cloned human DNA sequence, we have described in detail, in Sections 3.1–3.7, a protocol based on the method of Buckle and Craig *(1)* for ^{3}H-labeling of a DNA probe and its hybridization to replication-banded human chromosomes derived from blood lymphocytes. In addition, a number of alternative methods, which involve adaptation of the basic protocol for slightly different purposes, are briefly described in Section 3.8.

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

2. Materials

2.1. Labeling of Probe DNA

1. Tritiated nucleotides: Deoxy (1', 2', 5-^{3}H) cytidine 5'-triphosphate, ammonium salt (^{3}H-dCTP; Amersham code TRK625, approx 50 Ci/mmol); deoxy (1',2',2,8-^{3}H) adenosine 5'-triphosphate, ammonium salt (^{3}H-dATP; Amersham code TRK633, approx 60 Ci/mmol); (methyl, 1',2'-^{3}H) thymidine 5'-triphosphate, ammonium salt (^{3}H-dTTP; Amersham code TRK576, approx 120 Ci/mmol).
2. Dessicator and vacuum pump.
3. Multiprime labeling kit (Amersham code RPN1600 or Boehringer Mannheim code 1004 760 [for use with any labeled dNTP]) or nick-translation kit (Amersham code N.5500 [for use with any labeled dNTP]).
4. G50 medium Sephadex (Pharmacia): 5 g in 100 mL of pure water. Autoclave and store at room temperature.
5. Scintillation beta counter.
6. Scintillation fluid, 1 L: 5.5 g (2,5 diphenyloxazole PPO; Packard), 0.1 g POPOP (Packard) (1,4-bis[2-(5-phenyloxazolyl)]-benzene), 667 mL toluene, 333 mL Triton X-100.

   Stir in lightproof container in fume cupboard for 1–2 h. Store in brown bottle at room temperature.
7. Lyophilizer (e.g., Dynavac freeze-drying unit).

2.2. Chromosome Preparation

1. Whole blood (2–5 mL) collected in a sterile heparinized tube.
2. Tissue-culture-quality plastic centrifuge tubes (10 mL).
3. Phytohemagglutinin (PHA) (Wellcome): Reconstitute in 5 mL of distilled water as directed. Store 4°C for 2–3 wk only.
4. Culture medium 199 (e.g., Flow, GIBCO):

medium 199 (with Hanks' salts containing bicarbonate)	100 mL
Fetal calf serum	16 mL
L-Glutamine, 200 m*M*	0.3 mL
penicillin	10,000 U
streptomycin	10 mg
PHA	2.4 mL
HEPES 1*M*	2 mL

5. 5-Bromo-2-deoxyuridine (BrdU) (Sigma): Prepare a stock solution of 10 mg/mL in water; millipore filter (0.22 μ*M*). Dispense in 1–2-mL aliquots. Store wrapped in foil at –20°C. Use at 200 μg/mL (i.e., 0.1 mL

of 10 mg/mL stock/5 mL of blood culture). This chemical is mutagenic; therefore, wear gloves and do not inhale powder.

6. Thymidine (Sigma [cell-culture reagent]): Prepare a $10^{-3}M$ stock solution in pure water (i.e., 0.25 mg/mL); millipore filter. Store in 2-mL aliquots at –20°C. Use at a final concentration of $10^{-5}M$ in culture medium.
7. Phosphate-buffered saline—Dulbecco (PBS), pH 7.2 (calcium- and magnesium-free). Autoclave to sterilize.
8. Colchicine (BDH) or Colcemid® (Sigma or GIBCO): Prepare a sterile stock of 20 µg/mL in water. Store wrapped in foil in 1-mL aliquots at –20°C (light- and temperature-sensitive; will deteriorate over several years). For use, dilute 1/50 to give a final concentration of 0.4 µg/mL in culture medium.
9. 0.56% potassium chloride in water. Store at room temperature.
10. Cell fixative: 3/1 (v/v) analar methanol/glacial acetic acid. Prepare fresh for each harvest.
11. Glass dishes and slide carriers (Wheaton Z10-396-9): Wheaton horizontal staining dish with dish cover, removable glass slide rack with handle (10-slide capacity).
12. Good quality glass slides with frosted end and 22-mm × 60-mm cover slips: Soak in detergent (2% Decon®90 or Lipsol) overnight, rinse well in distilled water, soak in 1% HCl solution for at least 1 h, rinse well in distilled water, soak in two rinses of ethanol, and air-dry. Store in dust-free container until used.
13. Lightproof slide box (e.g., Kartell-Milano) to hold 25, 50, or 100 slides: Include a perforated tube containing silica gel as a desiccant.
14. Darkened alcove: Reserve a corner of the laboratory that can be isolated with a heavy floor-to-ceiling curtain. Facilities in the alcove should include a work bench, benchtop centrifuge (200*g*), water bath, and a table lamp with yellow-green darkroom lightbulb (Phillips, 240-250V PF710 B).

2.3. Slide Treatment Prior to Hybridization

1. RNase (Boehringer Mannheim 109142 or Sigma R5500): Prepare a 1 mg/mL stock solution in TE buffer (10 m*M* Tris-HCl; 1 m*M* EDTA, pH 7.4). Boil for 10 min to remove DNase contamination. Dilute to 100 µg/mL in 2× SSC; millipore filter. Store in 2-mL aliquots at –20°C (stable to freeze-thawing).
2. Ethanol (AR): Make up stocks of 10, 50, 75, and 95% ethanol in water and 100% ethanol.

3. SSC (20×) solution: sodium chloride 175 g, tri-sodium citrate 88 g. Make up to 1 L with distilled water. Store at room temperature.
4. Formamide (Merck AR): For each experiment, freshly deionize 200 mL of formamide by adding 20 g of Bio-Rad mixed-bed resin (AG®501-X8[D]), and then stir at room temperature for at least 2 h or overnight. Store at 4°C in a dark bottle. Formamide is toxic; therefore, wear gloves and handle in a fume hood.
5. DNA-denaturation solution: Deionized formamide, 140 mL; 0.25*M* EDTA, 0.1 mL, 2× SSC, 60 mL. Adjust to pH 7.0 with 5*M* HCl. Make up fresh prior to use.
6. Moist chamber: Any air-tight box with a sheet of Whatman filter paper moistened with 2× SSC will serve this purpose.
7. Black plastic lightproof bags to protect slides during various procedures.

2.4. Hybridization and Washing

1. 0.2% sodium dodecyl sulfate (SDS), 10 m*M* EDTA: Store at room temperature.
2. Dextran sulfate (Pharmacia): Prepare a 50% stock solution in distilled water. Boil to dissolve. Cool slowly. Store in 1-mL aliquots at –20°C. Stable to freeze/thawing.
3. Salmon-sperm DNA (Boehringer Mannheim): Prepare a 10 mg/mL solution. Cut threads of dehydrated DNA into small pieces with clean scissors. Add 500 mg to 50 mL of distilled water and dissolve in a 65°C water bath for several hours. Shear by passing through three successive hypodermic needles (19, 21, and then 25 gage) using a 50-mL syringe. Boil for 10 min. Store at –20°C.
4. 20× SSPE solution: NaCl, 17.4 g; $NaH_2PO_4 \bullet 2H_2O$, 2.8 g. Dissolve in 80 mL of distilled water. Add 10 mL of 200 m*M* EDTA (final conc. 20 m*M*). Adjust to pH 7.4 with conc. NaOH and make up to a final vol of 100 mL with distilled water. Store at room temperature.
5. Hybridization buffer:

Hybridization buffer:	1 mL	Final conc.
deionized formamide (*see* Section 2.3.4)	500 µL	50%
50% dextran sulfate	200 µL	10%
salmon-sperm DNA (10 mg/mL)	20 µL	0.2 mg/mL
20× SSPE	250 µL	5×
distilled water	30 µL	

   Prepare fresh for each experiment.
6. Cover-slip rubber-seal solution (e.g., Earth brand rubber solution for tube repairing).
7. Dry incubator or water bath, 42°C.

2.5. Autoradiography

1. Emulsion (Ilford nuclear emulsion L4 [crystal diameter 0.13 µm], Amersham RPN 41 [crystal diameter, 0.13 µm] or Kodak™ NTB-2 [crystal diameter 0.26 µm]). Prepare according to the directions of the manufacturer, in total darkness, and use the appropriate safelight only when necessary. Store in a glass jar in a lightproof box at 4°C (do not freeze). Prepare a 1/1 solution of L4 emulsion by melting, for example, 50 g of mousse at 46°C and mixing with 50 mL of pure water at 46°C (avoid forming air bubbles). Stir gently with a glass rod or plastic spoon (avoid contact of emulsion with any metal, to minimize formation of background latent image). Stand for 10 min at 46°C. Dispense into smaller aliquots in slide mailers, enough for an average-size experiment, or store whole at 4°C and remelt for multiple use until background becomes unacceptable. (It is important to ensure complete equilibration and melting of emulsion at 46°C, to avoid a heavy granular appearance on slides).
2. Wheaton glass dishes (3) and slide holder.
3. D19 developer (also available commercially): Metol (*p*-methylaminophenol sulfate) (BDH) 2 g; sodium sulfite (anhydrous) 90 g; sodium carbonate (anhydrous), 45 g; potassium bromide, 5 g; Quinol (hydroquinone) (BDH), 8 g. Dissolve in the order shown above and make up to 1 L with distilled water. Store in a brown bottle in the dark at room temperature. Remove an aliquot for each experiment and discard after each use. Wear gloves.
4. Stop solution: 1% glacial acetic acid in water.
5. Fixative: 200 mL of Ilford Hypam, 800 mL of distilled water. While stirring constantly, slowly add 25 mL of Ilford Hypam hardener. Store in a brown glass bottle in the dark at room temperature. (The hardener is necessary to prevent the emulsion from swelling and lifting off slides.) The fixative can be reused for several experiments. Wear gloves.
6. Darkroom: This should be equipped with a 46°C water bath and a safelight filter (Ilford 904 [dark brown] is suitable for Ilford L4 and Amersham RPN41 emulsion, or use a red safelight for Kodak™ emulsions).

2.6. Chromosome Staining

1. Hoechst 33258 (Sigma No. B.2883) Bisbenzimide trihydrochloride: Prepare stock solution of 1 mg/mL in distilled water. Aliquot into 2-mL vol; wrap in foil. Store at –20°C. This can be thawed and refrozen several times, and is light-sensitive. Avoid inhaling or contact with skin.
2. Phosphate-buffered saline (PBS) (BDH): Gurr® buffer tablets, pH 6.8.

3. Giemsa Gurr®'s improved R66 (BDH): Immediately before use, dilute Giemsa to 10% in PBS, pH 6.8. Filter into a coplin staining jar.
4. Table lamp with a long-wavelength UV light tube (Sylvania; Blacklite-blue, 15 W).

2.7. Chromosome Analysis

1. High-resolution light microscope fitted with 12.5× eyepieces and 10× and 100× objectives; green filter (546 nm) or blue-green interference filter.
2. Films for black-and-white photography: Agfa Copex™ rapid AHU TRI 13 gives good results.

2.8. Biotin Labeling of Probe DNA

1. Biotinylated dUTP (bio-11-dUTP, biotin-labeled deoxyuridine triphosphate) (TTP analog) (ENZO Biochem Inc., New York).
2. Proteinase K: 10 µL of proteinase K stock (2.5 mg/mL), 10 mL of 1*M* Tris-HCl, pH 7.4, 20 mL of 50 m*M* $CaCl_2$, 470 mL of distilled water. Store frozen in aliquots. Preincubate for 2 h at 37°C immediately prior to use to eliminate any contaminating proteins, such as nucleases.
3. Hybridization buffer:

Hybridization buffer:	1 mL	Final Conc.
deionized formamide (*see* Section 2.3.4)	500 µL	50%
50% dextran sulfate	200 µL	10%
20× SSC	100 µL	2×
salmon-sperm DNA (10 mg/mL)	100 µL	1 mg/mL
distilled water	100 µL	

Prepare fresh for each experiment.

4. PN buffer: 189.4 mL of 0.5*M* Na_2HPO_4, 10.6 mL of 0.5*M* NaH_2PO_4, 799.5 mL of distilled water, 0.5 mL of NP40. Store at room temperature.
5. PN buffer with 5% nonfat dry milk: 5 g of nonfat milk (Carnation skim milk powder, 1% fat) in 100 mL of PN buffer. Filter. Store at –20°C in 10-mL aliquots.
6. Fluorescein avidin DCS (avidin-FITC) (Vector A-2011, cell-sorter grade). Aliquot and store at –20°C. Use at 5 µg/mL diluted in PN buffer, 5% milk.
7. Normal goat serum: Dilute to 5% in PN buffer. Store at –20°C.
8. Biotinylated goat antiavidin-D (Vector BA0300): Aliquot and store at –20°C. Use at 5 µg/mL diluted in PN buffer, 5% normal goat serum.
9. Propidium iodide (5 µg/mL) or 4',6diamidino-z-phenyl-indole dihydrochloride (DAPI; Sigma D-1388) (0.25 µg/mL) in "antifade" (*p*-phenylenediamine dihydrochloride): 50 mg of *p*-phenylenediamine dihydrochloride (Sigma P1519) in 5 mL of PBS. Adjust to pH 8.0 with

0.5*M* bicarbonate buffer, pH 9.0. Add 1 mL of antifade to 9 mL of glycerol; millipore filter (0.22 μm). Dissolve counterstain (propidium iodide at a concentration of 5 μg/mL or DAPI at a concentration of 0.25 μg/mL). Store at 4°C in the dark. (The solution darkens with time, but is stable for at least 6 mo).

10. Fluorescence microscope: Normal microscope equipped with appropriate filters and light source. A Zeiss epifluorescence condenser based on reflected light bright-field excitation gives good results. The light source is generally a high-pressure mercury source, or, for blue excitation (for use with fluorescein isothiocyanate [FITC]), a high-power halogen source. Filter: blue (for FITC) (ZEISS 487709), transmittance range 450–490 nm. Photograph with fast color film—e.g., Daylight Ektachrome™, 200–400 ASA.

3. Methods

3.1. ^{3}H-Labeling of Probe DNA

1. Label 100–200 ng of DNA by multipriming or nick translation using 50–100 μCi of a cocktail of ^{3}H-nucleotides (e.g., ~20 μCi ^{3}H-dCTP, ~20 μCi ^{3}H-dATP, and ~40 μCi ^{3}H-dTTP, based on the specific activities given in Section 2.1) according to instructions with labeling kits. ^{3}H-Nucleotides are supplied in 50% alcohol, so they have to be dried before use, using a dessicator and vacuum pump. For multipriming, incubate the reaction for 2–6 h at 37°C, or overnight at room temperature. For nick translation, incubate for 2–3 1/2 h at 15°C.
2. Remove unincorporated nucleotides by passing the reaction mixture through a G50 Sephadex column (made up in a Pasteur pipet) and collect 100-μL fractions. Add 2 μL from each fraction to an aliquot of scintillation fluid and count in a beta counter. Pool the fractions from the first peak of radioactivity, which will contain the labeled probe. Divide into aliquots suitable for one experiment, lyophilize, and store at –20°C. Incorporation of 20–50% and a specific activity of 3–5 × 10^8 dpm/μg probe DNA should be achieved.

For probe DNA, either whole plasmid or eluted insert can be used, although we have found that the former sometimes gives unacceptably high background for unknown reasons. We have also used a single ^{3}H-nucleotide with satisfactory results, but use of three labeled nucleotides offers higher specific activity and greater sensitivity. Four labeled nucleotides can also be used. Multipriming gives better incorporation than nick translation, especially when the probe is small or small amounts of DNA are to be labeled.

3.2. Preparation of Human Chromosomes

The replication G-banding protocol of Zabel et al. *(2)* is described here, since this procedure yields banded chromosome preparations of high quality.

Full safety precautions should be used when handling unfixed biological materials. The operator should wear protective gloves and work in a class II biohazard hood. All materials should be discarded into hypochlorite or autoclaved.

1. Add 0.2 mL of sterile heparinized whole blood to a 10-mL culture tube containing 5 mL of medium 199 (Section 2.2.4). Place tubes in a sloping rack and culture for 72 h at 37°C (0.2 mL of whole blood in one culture tube should generate 5–10 slides).
2. Add 0.1 mL of BrdU to the blood culture to give a final concentration of 200 µg/mL and incubate inside a black plastic bag at 37°C for a further 16–17 h. BrdU blocks DNA synthesis in the middle of the S phase and is therefore incorporated only into early replicating bands. After the addition of BrdU and in subsequent steps, protect the cultures in black plastic bags and work in the darkened alcove. In these steps, absolute darkness is not essential, but using a darkened alcove is important to minimize nicking the BrdU-incorporated DNA by light. Other workers who have omitted this precaution have produced less than satisfactory chromosome-banding results. This arrangement is more convenient than the use of a standard darkroom throughout the entire experiment.
3. Remove BrdU by washing twice: Centrifuge the cells at 200*g* for 5 min, discard the supernatant into hypochlorite, and resuspend the cells in medium prewarmed at 37°C. Repeat centrifugation and resuspend the cells in 5 mL of prewarmed medium 199 containing $10^{-5}M$ thymidine. (This releases the block in the cell cycle.) Reincubate at 37°C for 6–7 h to reinitiate DNA synthesis and allow mitosis to proceed.
4. Add 0.1 mL of colchicine to give a final concentration of 0.4 µg/mL and incubate for 15–30 min at 37°C. Although the cells are already synchronized by the BrdU, colchicine helps to disrupt the mitotic spindle and gives better chromosome spreading and morphology.
5. Remove the medium by centrifugation at 200*g* for 5 min and resuspend in 7 mL of 37°C prewarmed hypotonic solution (0.56% KCl). Incubate at 37°C for 15 min to achieve sufficient swelling of the cells.
6. Add 3 mL of cell fixative and mix thoroughly. Centrifuge the cells for 10 min, and discard supernatant. Resuspend the pellet and slowly add a further 7 mL of cold (4°C) cell fixative (mixing well throughout). Stand

it on ice for 20 min before centrifuging, and repeat the fixative wash three times. At this stage, cells can be stored in fixative (wrapped in foil) at –20°C for a few days. If stored, the fixative should be changed before the slides are made.

7. After the final fixative wash, resuspend the cells in a small volume (e.g., 0.5 mL) of cold fixative and put 2–3 drops of cold cell suspension onto a sloping dry glass slide. Air-dry. Check the slide under a phase-contrast microscope and adjust the cell concentration accordingly. Store slides at –20°C in an air-tight, dark slide box containing desiccant, and seal with black tape. Slides should be stored for a few days before use and can be kept in this manner for at least 6 mo.

3.3. Treatment of Chromosome Spreads Prior to Hybridization

Perform in darkened alcove.

1. Equilibrate the slide box to room temperature before opening to avoid condensation on the slides.
2. Destroy cellular RNA (which causes nonspecific hybridization background around cells) by adding 80 µL of 100 µg/mL RNase to each slide. Apply a 22- × 50-mm cover slip. Incubate in a moist chamber at 37°C for 1 h.
3. Remove the cover slip carefully by vertically floating off in 2× SSC. Wash the slides in three changes of 2× SSC. Dehydrate for 30 s in each of 10, 50, 75, 95, and 100% ethanol. Air-dry.
4. Denature the chromosomal DNA into single strands by incubating the slides for 4 min at 65°C in prewarmed denaturation solution (*see* Section 2.3.5), or at 70°C for 2 min.
5. Immediately plunge the slides into 75% ethanol precooled at 4°C. Dehydrate in 95% ethanol, followed by absolute alcohol at room temperature. At this stage, slides can be stored in desiccant at 4°C for a few hours if necessary.

3.4. Hybridization and Washing

The quantities used are suitable for 10 slides (5 ng of labeled DNA/slide).

1. Add 5 µL of 0.2% SDS, 10 m*M* EDTA, and 300 µL of hybridization buffer to 50 ng of ^{3}H-labeled, lyophilized probe DNA (*see* Section 3.1). Mix well. Boil for 5 min to denature the probe. Chill on ice immediately.

Perform the following steps in darkened alcove.

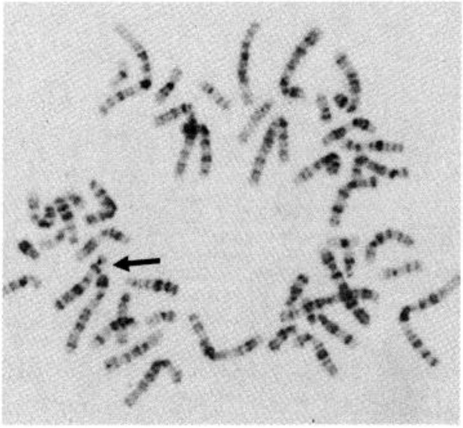


Fig. 1. Hybridization of a unique sequence (pyruvate dehydrogenase) to the p22.1–22.2 region of the human X chromosome *(3).* The arrow indicates the presence of a silver grain on the region. The washing stringency was 0.2× SSC at 60°C. The specific activity of the probe was 3×10^8 dpm/μg DNA. The probe concentration was 10–20 ng/mL (i.e., approx 0.5–1 ng/slide). The exposure time was 4 wk. Reprinted from ref. *6,* by permission of publisher.

2. Add 30 μL of hybridization/probe mix to each slide. With forceps, carefully lower a 22- × 50-mm cover slip from one end and avoid trapping any air bubbles.
3. Seal the edge of the cover slip with rubber solution. Place slides inside a sealed moist chamber on a dampened filter paper. Wrap the chamber in black plastic and hybridize overnight in a 42°C dry oven (or float in a water bath).
4. To wash the slides, peel off the rubber seal with forceps, and carefully remove the cover slip by vertically floating off in 5× SSC. Wash in 5× SSC for 5 min. Irrespective of the stringency requirements *(see below)* first wash slides in three changes of 2× SSC at room temperature for 20 min each to remove excess probe.
5. For the detection of chromosomal sequences highly homologous to the probe (i.e., high stringency conditions), use a subsequent wash of 0.5× SSC at 60°C for 1/2 h, followed by two changes of 0.1× SSC at 60°C for 1/2 h each (Fig. 1 and refs. *4* and *5*). For the detection of related sequences in a gene family (Fig. 2), use a lower-stringency wash of 1× SSC at 60°C for 1/2 h followed by 0.5× SSC at 60°C for two 1/2-h washes.

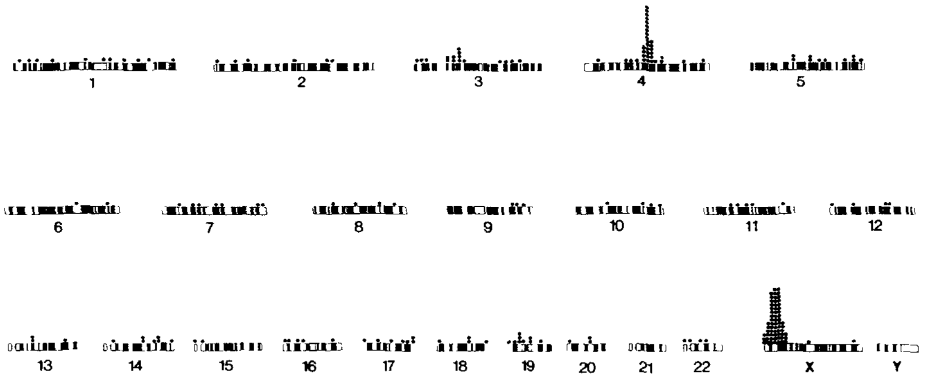


Fig. 2. Cumulative ideogram of silver-grain counts using a unique sequence probe (as in Fig. 1); 90 metaphases were analyzed. The use of a slightly reduced washing stringency, 0.5× SSC at 60°C, allowed the detection of the homologous gene on the X chromosome as well as a related gene on chromosome 4 *(3,6)*. Reprinted from ref. *3*, by permission of publisher.

Figure 3 shows the effect of a progressive increase in the specificity of a repetitive DNA probe for its cognate chromosome with increasing washing stringency.

6. Equilibrate the slides to room temperature in the same SSC concentration as the last wash for 15–60 min. Dehydrate in ethanol as in Section 3.3, Step 3. Store with desiccant at room temperature until ready for dipping in emulsion.

It is important to preheat all washing solutions in a water bath to the selected temperature to ensure attainment of the desired stringency. The chromosome preparations will withstand a temperature of up to 65°C for 1 h if greater stringency is necessary.

An alternative washing schedule using a solution containing formamide can also be used (e.g., ref. 7), but this involves the handling of a toxic chemical and is therefore less desirable.

3.5. Autoradiography

Perform in total darkness in the darkroom and use the prescribed safelight only when necessary.

1. Set the water bath at 46°C.
2. Remelt 1/1 diluted L4 emulsion in a 46°C water bath for 15–30 min. Stir gently with a plastic spoon (avoid bubbles). Stand for 10 min at 46°C to allow air bubbles to disperse.
3. Dip a control (blank, clean) slide by slowly lowering in and out of the emulsion. Stand it vertically to drain and check the evenness of emulsion under the safelight. Dip hybridized slides individually and stand them in a vertical position in a lightproof box for about 1 h, until hardened and dry.
4. Transfer the slides to a lightproof slide box containing desiccant, seal with tape, and leave it to expose at 4°C for an appropriate length of time (*see* Section 3.7).
5. To develop slides, equilibrate the slide box to room temperature before opening (to avoid condensation on slides, which affects the latent image). Develop slides in D19 developer diluted 1/1 in distilled water for exactly 5 min at 20°C (without agitation). Rinse in 1% acetic acid stop solution for 30 s, and fix for 10 min in Ilford Hypam fixative with hardener. Wash gently in slow-running tap water for 1 h and proceed to staining. (The blank slide should be included to serve as a control for the presence of background grains in the emulsion).

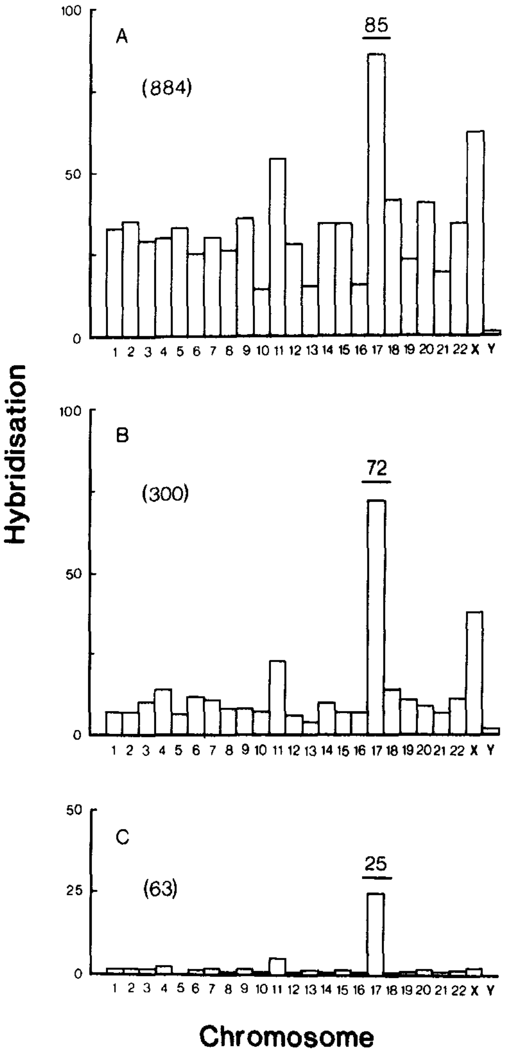


Fig. 3. Histograms of *in situ* hybridization of a chromosome 17 centromeric alpha satellite probe to human male chromosomes under different washing conditions: **(A)** 0.5x SSC; **(B)** 0.1x SSC; **(C)** 0.02x SSC; all at 60°C. Hybridization is expressed as the number of grains scored at the centromeric region of each chromosome. The values underlined are the number of grains on chromosome 17, whereas values in parentheses are the total number of grains on centromeres of all the chromosomes. At low stringency, the probe hybridizes to the alphoid subfamilies on all the chromosomes, whereas, with increasing stringency, the relative specificity of the probe increases from 9.6% (85/884) to 24% (72/300) to 39.7% (25/63) for chromosome 17 *(8)*. The specific activity was approx 10^8 dpm/µg DNA. The exposure time was 1 d. Reprinted from ref. *8*, by permission of publisher.

3.6. Chromosome Staining

Do not allow slides to dry during the staining procedure.

1. Stain slides in Hoechst 33258 diluted to 15 µg/mL in 2× SSC for 30 min. Rinse in 2× SSC.
2. Submerge slides horizontally and completely in a flat dish by covering with 2× SSC to a depth of 2–3 mm. Expose the slides for 1 h to a long-wave UV light by placing the light source 20 cm above the SSC level. (Avoid using excessive 2× SSC, since the increased depth will block penetration of the UV light.)
3. Rinse the slides in PBS, pH 6.8.
4. Stain vertically for 20–30 min in 10% Giemsa.
5. Rinse in PBS, pH 6.8, and air-dry vertically.

3.7. Chromosome Analysis

A blue-green interference filter or a green filter allows good visualization of the banding pattern, but the silver grains are more easily identified in white light without a filter.

1. Check the background grains on the blank slide.
2. Analyze the distribution of silver grains over the whole karyotype to identify the chromosome or chromosomes that give a significant accumulation of probe signal. This involves finding cells with well-spread chromosomes and recording individual grains (Fig. 1) on a cumulative ideogram (Fig. 2). Only grains on or touching a chromosome should be counted. Once a positive chromosome is identified, more grains can be scored to achieve a precise subregional localization using less-well-spread metaphases, as long as the essential chromosome is clear.

For a tandemly repeated probe or a mixture of unique-sequence probes, many grains may be present at one site (Figs. 4 and 5), so a good signal should be seen on most cells. Vary the exposure time (e.g., overnight to many days) to give an optimal signal to permit either grain counting (Fig. 3) or direct visualization and photographic presentation of results (Figs. 4 and 5). For a single-copy sequence, many cells will need to be scored to accumulate a significant signal, and a much longer exposure time of 10–30 d may be required *(9)*.

3.8. Alternative Methods

3.8.1. Use of Nonradioactive, Biotinylated Probe

This eliminates the handling of radioactive isotopes and allows results to be obtained more rapidly. The basic principles are the same as those described above, except that the DNA probe is labeled by substituting thymidine residues with biotinylated dUTP, which can then be detected by a series

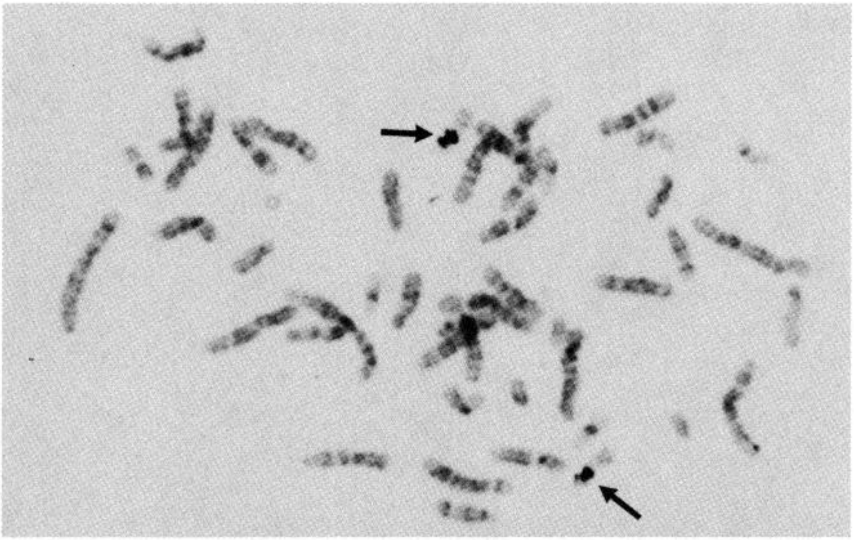


Fig. 4. Localization of a repetitive alpha DNA probe (as in Fig. 3) to the centromeric region of chromosome 17 (arrows). Extended autoradiographic exposure (3–4 d) has obliterated individual grains into a "blob," which provides a good visual picture, but one not useful for quantitative scoring *(8)*. Washing stringency was 0.1× SSC, 60°C. Reprinted from ref. *8*, by permission of publisher.

of fluorescence- or enzyme-conjugated antibiotin or antiavidin antibodies. This approach has been successfully used for the mapping of single-copy genes *(10–13)*. It is also increasingly being used for rapid sex determination and aneuploidy detection in interphase cell nuclei *(14,15)* and for direct analysis of chromosomal aberrations in metaphase cells *(16,17)*. In these applications, chromosome-specific repetitive probes *(15,18)*, or a mixture of "unique" sequences *(19)*, are used to "decorate" or "paint" specific regions of a chromosome. The availability of different fluorescent dyes allows multiple probes to be simultaneously applied to a slide.

The procedure described below is based on the use of a Y-specific repetitive sequence.

1. Label 1 µg of DNA by nick translation or by multipriming with biotin-dUTP according to the instructions of the supplier. Add 1 µCi of ^{3}H-dATP to trace the amount of biotin incorporation. Separate through a G50 Sephadex column (as in Section 3.1). Approximately 20% of the thymidine nucleotides in the probe DNA should be replaced with the biotinylated dUTP. Lyophilize the labeled probe and store at –20°C.

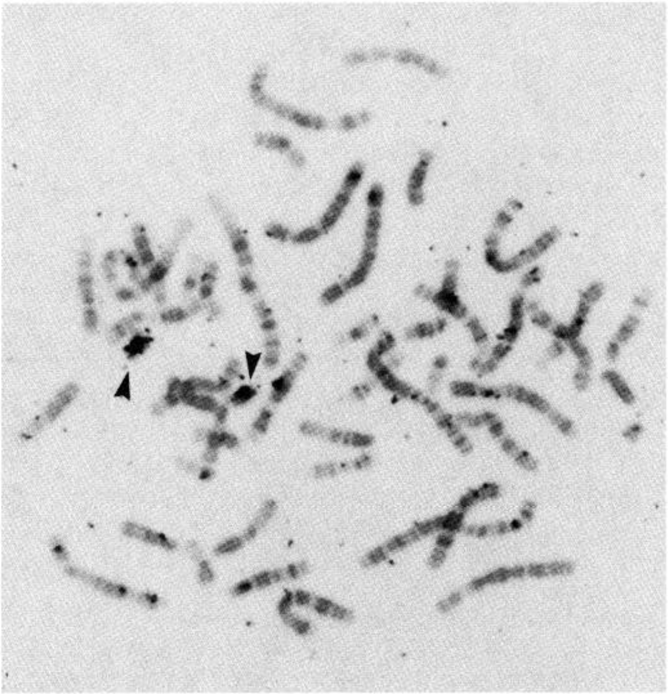


Fig. 5. *In situ* hybridization using a mixture of 50 chromosome-21 unique sequences (totaling approx 80 kb). Labeling was done with ^3H. The arrows indicate positions of the two chromosomes 21, which have been extensively "decorated" by silver grains *(9)*. The specific activity of probe was 8×10^8 dpm/μg DNA. Labeled DNA (10–20 ng) was used in each slide. The washing stringency was 0.1× SSC at 60°C for 1 h. The exposure time was 4 wk. Reprinted from ref. *9,* by permission of publisher.

2. Prepare nonbanded human chromosomes by standard cytogenetic methods from 72 h, PHA-stimulated blood-cell cultures. Add 0.4 μg/mL of colchicine for 1/2–2 h prior to harvest, to arrest the cells in mitosis. Make slides from the harvested cells after fixation in methanol/acetic acid (3/1), as described earlier.
3. Destroy cellular RNA and denature chromosomal DNA as in Steps 1–5 in Section 3.3. To improve the accessibility of the probe and detector molecules to the target DNA, after RNase treatment, incubate the slides in proteinase K solution for 7–8 min at 37°C. Rinse in 2 changes of 2× SSC and dehydrate (*see* Section 3.3, Step 3).

4. For 10 slides, add 5 µL of 0.2% SDS/10 m*M* EDTA and 300 µL of hybridization buffer to 300 ng of labeled (lyophilized) DNA. Mix well; boil for 5 min. Chill on ice immediately.
5. Add 30 µL of hybridization/probe mix to each slide. Subsequent steps are as in Section 3.4, Steps 2–5.
6. After the last wash in SSC, place slides immediately into PN buffer until ready to stain. To avoid nonspecific binding of the biotin detection reagents, do not allow the slides to dry.
7. Blocking (to decrease nonspecific binding of avidin): Blot excess PN buffer from the edge of the slide. Add 70 µL of PN buffer/5% milk. Apply a cover slip and leave at room temperature for 5 min.
8. Avidin–FITC incubation: Remove the cover slip by carefully sliding it off the slide. Drain the slide. Add 70 µL avidin–FITC (5 µg/mL in PN buffer/milk). Add a cover slip. Incubate for 15 min at 37°C in a moist chamber.
9. Washing: Prepare three coplin jars of PN buffer at 45°C in water bath. Wash slides at 45°C in three changes of PN buffer for 2 min each. Agitate occasionally. Slides can now be viewed for fluorescent signal. If the signal is visible, proceed to staining, Step 14. If the signal requires amplification, continue from Step 10.
10. Amplification: Drain slides. Add 70 µL of PN buffer, 5% normal goat serum. Add a cover slip. Leave at room temperature for 5 min. Remove the cover slip. Drain. Add 70 µL biotinylated goat antiavidin (5 µg/mL in PN buffer/goat serum). Add a cover slip. Incubate at 37°C for 20 min in a moist chamber.
11. Wash in three changes of PN buffer at 45°C for 2 min each.
12. Avidin–FITC incubation: Drain. Add 70 µL of avidin–FITC (5 µg/mL in PN buffer). Add a cover slip. Incubate at 37°C for 15 min in a moist chamber.
13. Wash in three changes of PN buffer at 45°C for 2 min each.
14. Staining: Drain well (but do not allow to dry). Add 20 µL of antifade/propidium iodide (or antifade/DAPI). Add a cover slip. (Only a very thin layer of the antifade/propidium iodide must be used).
15. View under a fluorescence microscope (Fig. 6). If the signal needs more amplification, repeat steps 10–14 after washing off the cover slip in PN buffer.

The fluorescein and propidium iodide are excited at 450–490 nm, so the hybridized areas appear as yellow-green spots on red chromosomes or cell nuclei. If DAPI counterstain is used, it is excited independently to

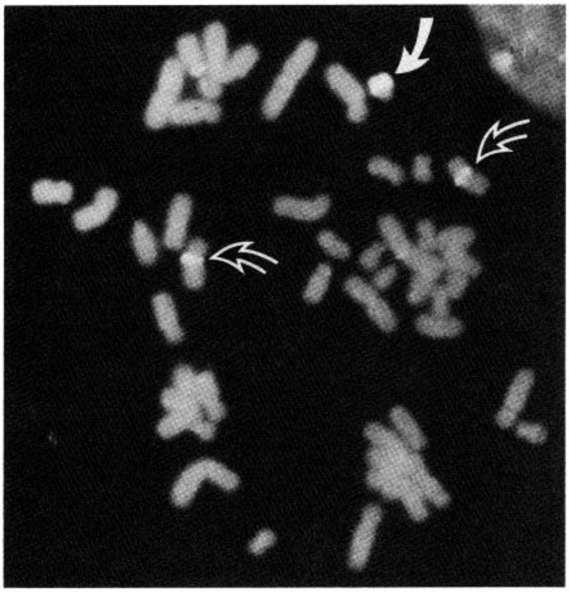


Fig. 6. Fluorescence hybridization *in situ* using a chromosome-Y repetitive probe. The probe hybridizes strongly to the q12 heterochromatic region of the Y chromosome (solid arrow), and less strongly to the q12 heterochromatic region of chromosome 9 (open arrows).

fluorescein and is viewed with UV excitation. Each field can be observed for several minutes, but some fading does occur, so photography is recommended.

To store slides, remove the cover slip in PN buffer, and rinse several times in PN buffer. Stained slides can be reexamined or restained, and can be stored for several weeks in PN buffer at 4°C in the dark.

3.8.2. Other Labeling and Detection Systems

1. Other detection systems for visualizing biotinylated probes include the use of alkaline phosphatase or peroxidase-labeled antisera, which form dark crystals of reaction products in the presence of appropriate substrates. This signal can be further amplified by the use of immunogold and silver precipitation, and can be used in conjunction with replication banding or Giemsa staining of the chromosomes *(11)*.
2. ^{125}I-labeled nucleotide may be used to label the probe. This is no longer commonly used, because of its greater radiation hazard and its relatively short half-life of 60 d. Its higher specific activity offers a stronger signal, but at the same time produces a higher background.

3.8.3. Nonreplication G-Banding of Chromosomes

In some situations, live cells are not available for replication banding of the chromosomes (e.g., cells without BrdU incorporation stored in fixative after routine cytogenetic analysis). Use of these cells necessitates a series of steps:

1. G-banding of chromosomes,
2. Prephotographing selected metaphases,
3. Destaining,
4. Hybridization with labeled probe and autoradiography,
5. Counterstaining with Giemsa, and
6. Rephotographing the same cells to allow silver grains to be localized to bands on relevant chromosomes.

This method is very tedious and accurate localization can be difficult. The efficiency of hybridization may also be affected by the pretreatment of the chromosomes before hybridization.

3.8.4. Metaphase Preparations of Different Cell Types

With very minimal modifications, the replication banding method described in Section 3.2 can be used for cultured fibroblasts, somatic cell hybrids, continuous lymphoid cell lines, and other cell lines.

3.8.5. Interphase Nuclei Preparations

In situ hybridization of interphase cells (Fig. 7) promises to be a rapid method that will increasingly be used to detect aneuploidy. Prepare interphase nuclei from whole blood by separating the nucleated cells through Ficoll/Hypaque. Wash. Cytospin the cells onto a glass slide. Fix in cold acetone, methanol/acetic acid (3/1), or Carnoy's fluid (ethanol/chloroform/glacial acetic acid, 6/3/1) for 10 min. Interphase nuclei for amniotic fluid cells, chorionic villi aspirates, buccal smears, sperm samples, and other fresh tissue samples can also be prepared *(14)*.

3.8.6. Replication Banding of Mouse Chromosomes

Since the studies of human and mouse genetics very often complement each other, mapping DNA sequences onto mouse chromosomes is an important application of the *in situ* technique. A method suitable for this purpose is outlined below *(20)*.

Kill a mouse and dissect out the spleen under aseptic conditions. Cut up the spleen and pass through a sieve into medium. Set up cultures from these spleen lymphocytes at 4×10^6 cells/mL (one spleen has approx 10^8 cells) in RPMI 1640/10% FCS/HEPES/antibiotics and concanavalin A (Sigma C5275)

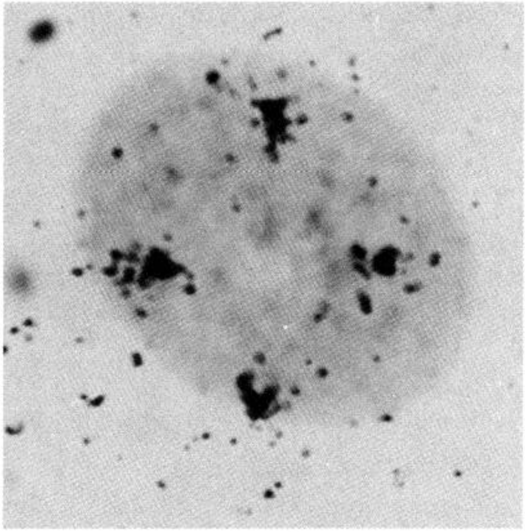


Fig. 7. Detection of aneuploidy on an interphase cell nucleus derived from a patient with four X chromosomes. These four chromosomes are clearly identified as four foci of signals using a chromosome-X-specific alpha-repetitive probe *(8)*.

at 3 µg/mL to induce mitotic activity. After culturing for 2–3 d at 37°C, add 400 µg/mL of BrdU and continue incubation for a further 17 h. Wash cells as in Section 3.2, Step 3 and incubate for a further 4–5 h at 37°C in the presence of $10^{-5} M$ thymidine. Add Colcemid® (0.5 µg/mL) 15–30 min before harvesting as described in Sections 3.2, Steps 5–7. All procedures after the addition of BrdU should be carried out in the darkened alcove.

Acknowledgments

We thank Ian and Sally Craig for advice on the replication-banding procedure, Graham Webb for advice on preparation of mouse chromosomes, and Academic Press, New York Academy of Sciences, Mary Ann Liebert Inc., and Springer-Verlag for permission to reproduce published figures. KHC is a Senior Research Fellow of the National Health and Medical Research Council of Australia.

References

1. Buckle, V. J. and Craig, I. W. (1986) In situ hybridisation, in *Human Genetic Diseases: A Practical Approach.* Davies, K., ed. IRL, Oxford, pp. 85–100.
2. Zabel, B. U., Naylor, S. L., Sakaguchi, A. Y., Bell, G. I., and Shows, T. B. (1983) High resolution chromosomal localisation of human genes for amylase, proopiomelano-

cortin, somatostatin, and a DNA fragment (D3S1) by in situ hybridisation. *Proc. Natl. Acad Sci USA* **80,** 6932–6936.

3. Brown, R. M., Dahl, H.-H. M , and Brown, G. K. (1989) X-Chromosome localisation of the functional gene for the El alpha subunit of the human pyruvate dehydrogenase complex. *Genomics* **4,** 174–181

4 Choo, K. H., Vissel, B , Brown, R , Filby, R. G., and Earle, E. (1988) Homologous alpha satellite sequences on human acrocentric chromosomes with selectivity for chromosomes 13, 14 and 21: Implications for recombination between nonhomologues and Robertsonian translocations. *Nucleic Acids Res.* **16,** 1273–1284

5. Choo, K. H., Vissel, B., and Earle, E. (1989) Evolution of alpha satellite DNA on human acrocentric chromosomes *Genomics* **5,** 332–344.

6. Brown, G. K., Brown, R. M., Scholem, R. D., Kirby, D M , and Dahl, H-H. M. (1989) The clinical and biochemical spectrum of human pyruvate dehydrogenase deficiency. *Ann NY Acad. Sci.* **573,** 360–368.

7 Harper, M E. and Saunders, G F (1981) Localisation of single copy DNA sequences on G-banded human chromosomes by in situ hybridisation *Chromosoma* **83,** 431–439.

8 Choo, K. H , Brown, R , Webb, G., Craig, I., and Filby, G. (1987) Genomic organisation of human centromeric alpha satellite DNA· Characterisation of a chromosome 17 alpha satellite sequence *DNA* **6,** 297–305.

9 Choo, K H , Filby, G , Earle, E., and Brown, R. (1988) Isolation of human chromosome 21 sequences and their application to in situ hybridisation. *Hum Genet* **81,** 49–53.

10 Albertson, D. G., Fishpool, R , Sherrington, P., Nacheva, E., and Milstein, C (1988) Sensitive and high resolution in situ hybridisation to human chromosomes using biotin labeled probes Assignment of the human thymocyte CD1 antigen genes to chromosome 1. *EMBO J* **7,** 2801–2805

11 Bhatt, B , Burns, J , Flannery, D , and McGee, J O'D. (1988) Direct visualisation of single copy genes on banded metaphase chromosomes by nonisotopic in situ hybridisation. *Nucleic Acids Res.* **16,** 3951–3961.

12 Cherif, D., Bernard, O , and Berger, R. (1989) Detection of single-copy genes by nonisotopic in situ hybridisation on human chromosomes *Hum. Genet* **81,** 358–362.

13. Viegas-Pequignot, E , Dutrillaux, B , Magdelenat, H , and Coppey-Moisan, M (1989) Mapping of single-copy DNA sequences on human chromosomes by in situ hybridisation with biotinylated probes: Enhancement of detection sensitivity by intensified-fluorescence digital-imaging microscopy *Proc Natl. Acad. Sci USA* **86,** 582–586

14 Burns, J , Chan, V T. W , Jonasson, J A., Fleming, K. A , Taylor, S , and McGee, J O'D (1985) Sensitivity system for visualising biotinylated DNA probes hybridised in situ. Rapid sex determination of intact cells. *J. Clin Pathol.* **38,** 1085–1092.

15. Cremer, T , Landegent, J , Bruckner, A , Scholl, H. P., Schardin, M , Hager, H. D , Devilee, P., Pearson, P , and Van Der Ploeg, M. (1986) Detection of chromosome aberrations in the human interphase nucleus by visualisation of specific target DNAs with radioactive and nonradioactive in situ hybridisation techniques: Diagnosis of trisomy 18 with probe L1 84 *Hum Genet* **74,** 346–352.

16 Cremer, T , Lichter, P , Borden, J., Ward, D C , and Manuelidis, L (1988) Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridisation using chromosome-specific library probes *Hum. Genet.* **80,** 235–246

17 Lichter, P., Cremer, T., Tang, C. C., Watkins, P C., Manuelidis, L., and Ward, D. C

(1988) Rapid detection of human chromosome 21 aberrations by in situ hybridisation *Proc Natl. Acad. Sci. USA* **85**, 9664–9668.

18. Pinkel, D., Straume, T., and Gray, J W. (1986) Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridisation. *Proc. Natl. Acad Sci USA* **83**, 2934–2938.
19. Lichter, P., Cremer, T., Borden, J., Manuelidis, L., and Ward, D. C. (1988) Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridisation using recombinant DNA libraries. *Hum. Genet.* **80**, 224–234.
20. Webb, G. C., Lee, J. F., Campbell, H. D., and Young, I. G. (1989) Haemopoetic growth factor gene IL3 and IL4 mapped to the same locus on mouse chromosome 11 *Cytogenet Cell Genet.* **50**, 107–110

CHAPTER 22

DNA Fingerprinting Analysis

Methodology and Its Applications

Richard A. Wells and Swee Lay Thein

1. Introduction

In the early 1980s, a new type of DNA element was described that consisted of arrays of tandemly repeated short sequence units of 9–64 bp (reviewed in ref. *1*). A large number of these arrays have subsequently been discovered. These have become known as minisatellites (by analogy with much larger satellite sequences) *(2)*, and are also known as VNTRs (variable number tandem repeats) *(3)*. Many minisatellites display a high degree of allelic variation in the number of repeats in the array, which is detectable as length polymorphisms in Southern blots; these minisatellites are often referred to as HVRs (hypervariable regions). The degree of heterozygosity at HVR loci is high (>99% for some HVRs), and thus, such hypervariable minisatellite loci provide highly informative genetic markers ideal for linkage analysis in humans.

Minisatellites exist as "families," the members of which are related by similarity of the sequences in their repeated units. This similarity allows cloned DNA segments containing such minisatellites to detect alleles from multiple loci simultaneously when used as hybridization probes under conditions of relaxed stringency. The resultant complex pattern of hybridizing fragments constitutes an individual-specific "DNA fingerprint," the component alleles of which are somatically stable and are inherited in a Mendelian fashion *(4)*. DNA fingerprinting has proved to be an extremely valuable tool in determination of correct paternity, in the unambiguous identification of an indi-

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by· C. Mathew Copyright © 1991 The Humana Press Inc , Clifton, NJ

vidual for forensic purposes, in transplantation biology, and as a tool to scan a large number of dispersed loci simultaneously in the study of allelic loss in neoplastic tissue *(1)*. A large number of probes containing disparate repeats have been devised for DNA fingerprinting; in this chapter, the use of three of these, the myoglobin-derived 33.6 and 33.15 minisatellite probes *(2)*, and the 'phage M13 gene III tandem repeats *(5)*, is described.

Minisatellites that have the ability to detect multiple loci simultaneously are referred to as "polycore" probes to distinguish them from the hypervariable single-locus probes, e.g., the probe detecting the HVR 3' to α-globin gene cluster *(6)*. Furthermore, it has been shown that large DNA fragments could be isolated by molecular cloning from a human DNA fingerprint; under hybridization conditions of high stringency, these isolated minisatellite clones act as locus-specific probes that detect extremely variable minisatellites with heterozygosities ranging from 90 to 99% *(7,8)*. The extreme heterozygosity combined with their sensitivity in Southern blot hybridizations, even when not aided by enzymatic amplification of genomic DNA, makes them extremely useful in individual identification. Furthermore, a panel of these locus-specific probes can be pooled to generate multilocus Southern blot patterns to provide a "reconstituted" DNA fingerprint. However, the level of individual specificity attained is not as high as can be achieved using "polycore" probes.

2. Materials

1. The apparatus used for agarose gel electrophoresis and Southern transfer has been described in Chapter 15.
2. 2× TY medium: 16 g/L Bacto tryptone, 10 g/L yeast extract, 5 g/L NaCl.
3. Top agar: 10 g/L Bacto tryptone, 8 g/L NaCl, 8 g/L agar.
4. H agar: 10 g/L Bacto tryptone, 8 g/L NaCl, 12 g/L agar.
5. *Escherichia coli* JM101 is: *lacpro, thi, supE,* F' *tra*D36, *por*AB, *lac* 1 of ΔZM15, and must be maintained on glucose/minimal medium to select for the F' plasmid. Glucose/minimal medium: to make 1 L, add 200 mL 5× M9 salts and 20 mL 20% glucose to 780 mL sterile deionized water. The 5× M9 salts is made by dissolving the following salts in deionized H_2O to a final vol of 1 L: 64 g $Na_2HPO_4 \cdot 7H_2O$, 15 g KH_2PO_4, 2.5 g NaCl, and 5 g NH_4Cl. Aliquot the salt solution into 200 mL and autoclave.
6. Xgal: 2% in Dimethylformamide
7. IPG: Isopropylthio-β-D-galactoside. Make a solution of IPTG by dissolving 2 g of IPTG in 8 mL of distilled H_2O. Adjust vol to 10 mL with distilled H_2O and sterilize by filtration through a 0.22-μm disposable filter.
8. TE-saturated phenol.

9. 3*M* Sodium acetate, pH 5.5.
10. PEG/NaCl: 20% polyethylene glycol 6000, 2.5*M* NaCl (store at 4°C).
11. TE: 10 m*M* Tris-HCl, pH 8, 1 m*M* EDTA.
12. dNTP stocks: 0.5 m*M* dATP, dGTP, dTTP, dCTP, pH 7. AGT mix: Equal vol of TE and 0.5 μ*M* dATP, dGTP, dTTP.
13. 10× Klenow reaction buffer: 100 μ*M* Tris-HCl, pH 8, 50 m*M* $MgCl_2$.
14. 10× *Eco* RI/Hind III reaction buffer: 10 m*M* Tris-HCl, pH 8, 600 m*M* NaCl, 70 m*M* $MgCl_2$, 70 m*M* β-mercaptoethanol.
15. Alkali stop solution: 1.5*M* NaOH, 0.1*M* EDTA.
16. Agarose gel loading buffer: 0.25% Bromophenol blue, 0.25% xylene cyanol, 15% Ficoll type 400, in water.
17. TAE buffer: 40 m*M* Tris-HCl, 5 m*M* sodium acetate, 1 μ*M* EDTA, pH 7.7.
18. Heparin prehybridization buffer: 50% Formamide, 3× SSC, 0.2% SDS, 50 μg/mL heparin. Heparin hybridization buffer: 50% formamide, 3× SSC, 5% dextran sulfate, 0.2% SDS, 200 μg/mL heparin.
19. Competitor DNA. Prepare a 10-mL solution of human genomic DNA in TE at a concentration of 0.5 mg/mL. Shear by sonicating at 50 W for ten 15-s bursts, separated by 5-s intervals. Alternatively, the DNA can be sheared by drawing it vigorously through a narrow gage needle repeatedly. Check the resultant fragment size by running a 1-μL aliquot on a small agarose gel. The average fragment size should be 200–600 bp.
20. SSC (standard saline citrate): 150 m*M* NaCl, 15 m*M* sodium citrate, pH 7.
21. Nylon membrane for Southern transfer of DNA, e.g., Amersham Hybond-N.
22. Kit for random hexamer labeling of probes using ^{32}P, available from several different manufacturers.
23. Phosphate based buffer: 0.5*M* $NaHPO_4$, pH 7.2, 7% SDS, 1% BSA, 1 m*M* EDTA. Make 1*M* $NaHPO_4$, pH 7.2 stock from 71 g/L Na_2HPO_4 anhydrous, e.g., BDH 10249, and adjust pH to 7.2 with approx 4 mL/L 85% orthophosphoric acid.

3. Methods

3.1. Preparation of Single-Stranded M13 Template

The DNA fingerprinting probes described here are most effective when radiolabeled by primer extension. For this reason, the probes are maintained as single-stranded M13 constructs, from which fresh template may be prepared by a simple transformation procedure.

3.1.1. Preparation of Competent Cells

1. Inoculate 5 mL of 2× TY with a single colony of JM101 and grow to stationary phase by incubating with shaking at 37°C overnight.
2. Add 500 µL of the overnight culture to 50 mL of 2× TY. Incubate at 37°C with shaking until the OD_{550} of the culture is ~0.3.
3. Pellet the cells by centrifugation at 3000*g* for 5 min. Discard the supernatant.
4. Resuspend the cells with gentle agitation in 25 mL ice-cold 50 m*M* $CaCl_2$ and let stand on ice for 20 min.
5. Repellet the cells again at 3000*g* for 5 min. Discard supernatant and resuspend cells gently in 5 mL ice-cold 50 m*M* $CaCl_2$. These are competent after standing on ice for 2 h.

3.1.2. Transformation of Competent Cells with Single-Stranded M13 DNA

1. Dilute single-stranded DNA probe (M13 vector, 33.6, or 33.15) to 1 ng/µL in TE.
2. Pipet 400-µL aliquots of competent JM101 cells into four sterile 15-mL tubes. Add 1, 5, 10, and 50 µL of the diluted single-stranded DNA solution to each. Mix by gently flicking the tubes and let stand on ice for 1 h.
3. "Heat shock" by incubating the tubes at 42°C for 3 min, then place back on ice immediately.
4. To each tube, add 40 µL of XGal, 40 µL of IPTG, and 200 µL of JM101 from the overnight culture (*see* Section 3.1.1, Step 1). Add 3 mL of molten top agar (kept at 40–50°C in a water bath) and mix by inversion. Pour immediately onto an H-agar plate. Allow the top agar overlay to solidify for 5 min, then invert the plates and transfer to a 37°C incubator. After overnight incubation, plaques will be evident in a bacterial lawn. M13 vector alone containing no inserts should give blue plaques; 33.6 and 33.15 recombinants should give clear plaques.

3.1.3. Preparation of Single-Stranded Template

Single-stranded template DNA can be prepared from plaques in a rapid "miniprep" procedure.

1. Prepare a fresh 5-mL overnight culture of JM101 in 2× TY as before.
2. Dilute the overnight JM101 culture 1:100 in 2× TY. Dispense 1.5 mL of diluted culture into a sterile 15-mL tube for each plaque that is to be processed.

3. Inoculate each 1.5-mL culture with phage by stabbing an individual plaque with a sterile plastic microloop and then swirling the loop in the culture.
4. Incubate the cultures at 37°C with shaking for 5 h.
5. Transfer the cultures to microcentrifuge tubes and pellet in a microcentrifuge for 5 min. Transfer the supernatants to fresh microfuge tubes containing 200 µL of PEG/NaCl solution. Mix by inversion and leave to stand at room temperature for 15 min.
6. Centrifuge for 5 min. Discard supernatant and recentrifuge for a further 5 min. Carefully remove all traces of PEG with a drawn-out Pasteur pipet. Small white pellets should be visible at this stage.
7. To each tube, add 100 µL of TE and 50 µL of Tris-saturated phenol. Vortex for 15 s and let stand for 10 min. Vortex again for 10 s and centrifuge for 5 min.
8. Transfer each aqueous (upper) phase to a fresh tube. To each, add 10 µL of 3.0*M* sodium acetate, pH 5.5 and 300 µL of cold (–20°C) absolute ethanol. Mix thoroughly and precipitate overnight at –20°C.
9. Pellet the DNA by centrifuging for 15 min. Remove the supernatants and wash the pellets with 500 µL of cold (–20°C) absolute ethanol. Remove supernatants with a drawn-out Pasteur pipet and dry the pellets under vacuum for 3 min. Resuspend each pellet in 50 µL of TE.
10. This procedure ordinarily yields ~5 µg DNA. To assess the concentration and mol wt of the harvested DNA, electrophorese a 2-µL aliquot of each sample together with 100 ng of vector M13 DNA and 100 ng of the DNA used for transformation.

3.2. Radiolabeling Probes 33.6 and 33.15 for DNA Fingerprinting

1. To a microcentrifuge tube, add 400 ng of single-stranded template DNA, 4 ng of 17-mer sequencing primer, 1 µL of 10× Klenow polymerase reaction buffer, and sterile distilled water to a final vol of 10 µL. Incubate at 55–60°C for 45 min. Spin the tube briefly in a microcentrifuge. The annealed template can be stored at –20°C at this stage.
2. To the annealed template, add 10 µL of AGT mixture, 6 µL of TE, and 30 µCi of [α-^{32}P] dCTP. Add 6 U of Klenow polymerase and mix by pipeting gently up and down. Incubate at 37°C for 30 min.
3. Add 2.5 µL of 0.5 m*M* (cold) dCTP and continue incubation at 37°C for 15 min.
4. Add 3 µL of *Eco*RI/*Hind*III 10× restriction buffer, 3 µL of 10 m*M* sper-

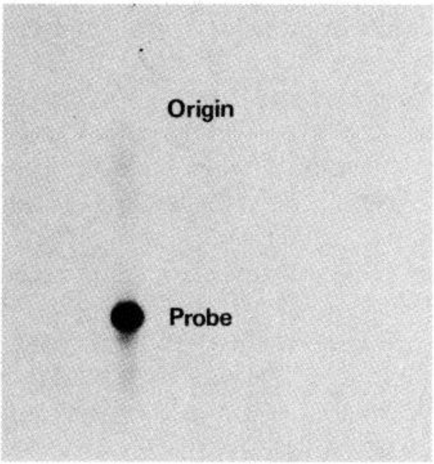


Fig. 1. Autoradiograph of a low-melting temperature agarose gel used for separation of the radiolabeled 33.6 minisatellite probe. The origin of the track and the strong probe band are shown.

midine trichloride, pH 7, and 15 U of the appropriate restriction enzyme [*Eco* RI for 33.6 and *Hind* III for 33.15]. Mix by pipeting up and down and incubate at 37°C for 45 min.

5. Stop reaction by adding 5.2 µL of alkali stop solution.
6. Add 5 µL of agarose gel loading buffer.
7. The labeled probe is purified by agarose electrophoresis. Prepare a 1.2% low-melting temperature agarose gel in TAE buffer.
8. Load and run the gel in a cold room at 7 V/cm for 2 h. The bromophenol blue dye front should have migrated ~6 cm from the origin.
9. Wearing gloves and goggles, slide the gel out of the gel tray and onto a sheet of cling film in an X-ray cassette. In a dark room, place a sheet of X-ray film over the gel. Develop the film after a 5-min exposure.
10. The autoradiograph should show a very strong probe band and weak background fogging, against which the gel slot should be visible (*see* Fig. 1). Measure the distance from the slot to the probe band and cut the corresponding slice from the gel using a scalpel blade. The probe migrates at about the same position as the bromophenol blue dye front, whereas unincorporated nucleotide gives a more diffuse signal further down the gel.

11. Transfer the agarose slice to a 5-mL plastic tube. Add 500 µL of sterile distilled H_2O and 10 µL of competitor DNA solution. Melt the probe at 100°C for 5 min.
12. To measure the specific activity of the probe, take a 10-µL aliquot and measure the radioactivity by scintillation counting. A standard labeling reaction using 400 ng of template should yield a total of $1–3 \times 10^7$ cpm.

3.3. Radiolabeling of M13 Vector for DNA Fingerprinting

The minisatellite region of M13 is located in 'phage gene III, distant from the cloning site of the vector. A modified procedure is used in preparing this probe for DNA fingerprinting.

1. Anneal 4 ng of 17-mer sequencing primer to 400 ng of single-stranded M13 vector as described for probes 33.6 and 33.15.
2. Spin briefly in a microcentrifuge. Add 10 µL of AGT mixture, 6 µL of TE, 30 µCi (3 µL) [α-^{32}P] dCTP, and 2 U of Klenow polymerase. Mix by pipeting gently up and down.
3. Incubate for only 15 min at 37°C, then stop the reaction by adding 70 µL of 3× SSC. Remove unincorporated label by spun column chromatography through G50-80 Sephadex (*see* Chapter 24). *Do not* boil this probe; it must remain partially double-stranded to work.

3.4. Radiolabeling a Panel of Locus-Specific HVR Probes

The Southern blot can be hybridized sequentially with a panel, usually five, of locus-specific HVR probes. Label 5–10 ng of each probe separately with ^{32}P-dCTP by the standard random priming method using a kit. Alternatively, a pool consisting of 3–5 ng of each probe could be made and the mixture radiolabeled. Labeled probes are purified by Sephadex G50–80 spun column, and counted in scintillation fluid.

3.5. Gel Electrophoresis and Filter-Transfer of DNA for Fingerprinting Analysis

Gel electrophoresis and Southern blot transfer have been described in detail in Chapter 15. However, the particular conditions that optimize the efficiency of DNA fingerprint analysis are outlined briefly here.

3.5.1. Restriction Analysis of DNA (see Note 2)

DNA obtained by the usual extraction procedures is of sufficiently high quality for fingerprinting analysis. To expose the variation in the number of

repeated units within the minisatellites so that maximum resolution can be obtained, it is necessary to digest the DNA with a restriction enzyme that cuts frequently but not within the repeated units of the HVRs. For the HVRs discussed here, *Hinf*I, *Alu* I, *Hae* III, and *Mbo* I are suitable.

Digest 5 µg DNA per gel track with the appropriate enzyme. In DNA fingerprinting analysis, not only the presence but also the intensity of a given allele is relevant; hence, it is important that the amount of DNA loaded is equivalent across all the tracks. DNA concentrations as well as completeness of digestion can be checked by running a small aliquot in a minigel.

3.5.2. Electrophoresis Conditions

Conditions are geared toward obtaining an optimum resolution of high-mol-wt bands between 4–25 kb. Generally, a 0.8–1.0% agarose gel run at 2 V/cm in TAE buffer with at least one change of buffer for every 12–16 h of electrophoresis is optimal. For fingerprinting analysis using "polycore" probes, electrophorese the 2.3-kb λ *Hind* III marker off the end of a 22-cm long gel; this involves a run of 48 h or more. For the locus-specific HVR probes, electrophorese until the 1-kb band is about 1 cm from the bottom of a (22-cm long) gel, which involves a run of 24–36 h.

3.5.3. DNA Transfer

After electrophoresis, the gel is stained with ethidium bromide and photographed on a short wave UV transilluminator to ensure that the samples have run far enough, and to record the position of the marker. The gel is also trimmed, if necessary.

1. Depurinate the gel in 0.25*M* HCl with gentle shaking for 8–10 min. Rinse the gel with distilled water.
2. Denature the gel in 0.5*M* NaOH/1.5*M* NaCl for 15 min twice with gentle agitation. Rinse the gel with distilled water.
3. Neutralize gel in 0.5*M* Tris-HCl 7.5/1.5*M* NaCl for 15 min twice.

The gel is now ready for DNA transfer by the usual method. DNA is fixed onto the nylon membrane by UV transillumination for 2 min, or baked for 1–2 h at 80°C.

3.6. Hybridization Conditions and Washing of Filters

DNA fingerprinting using the "polycore" probe depends on the cross-hybridization of many similar (but not identical) minisatellite sequences with a single probe. Thus, hybridization and posthybridization washing are performed under conditions of relaxed stringency, compared to those employed

Table 1
Washing Conditions for "Polycore" DNA Fingerprinting Probes

Probe	% SDS	× SSC	Temp	Time
M13	0.1	3 ×	Room temp	15–30 min × 2
33.6, 33.15	0.1	1 ×	61°C	15–30 min × 2

for detection of specific sequences using locus-specific HVR probes. For the "polycore" minisatellites, the heparin prehybridization and hybridization buffers allow this reduction of stringency without an unacceptable increase in nonspecific "background" probe binding. A phosphate-based buffer is suitable for use with the locus-specific probes.

Wet the filters in distilled water and then incubate in 1× SSC, 0.1% SDS at 37°C for 30 min. Seal them in a polythene pouch, two filters per pouch, each with the side to which DNA is bound facing outward.

3.6.1. "Polycore" DNA Probe

1. Add 15 mL of heparin prehybridization buffer that has been warmed to 37°C. Incubate the filters at 37°C for 3–24 h.
2. Warm 7.5 mL of heparin hybridization buffer to 37°C. Add probe ($4 \times 10^6 - 1 \times 10^7$ cpm) and mix by inversion.
3. Cut open one side of the bag containing the filters and squeeze out the prehybridization buffer. Add the hybridization buffer containing the probe and reseal the bag, taking care not to trap air bubbles. Spread the probe evenly using a roller, and incubate overnight at 37°C.
4. The optimum stringency of washing depends on the probe used, and is summarized in Table 1.
5. After washing, monitor the filters with a hand-held Geiger-Muller monitor. The level of bound radioactivity should be ~10–50× greater than that observed for unique sequence probes. Wrap the filters in cling film and autoradiograph between two intensifying screens at –70°C, as for ordinary hybridizations (*see* Note 1).
6. Develop the film after an overnight exposure.
7. If there is a high degree of background signal, rewash the filters for 30 min at 65°C in 1× SSC/0.1% SDS and repeat the autoradiography.
8. For maximum resolution, a 5–10-d exposure with no intensifying screens may be necessary.

3.6.2. Locus-Specific HVR Probes

1. Add 8 mL of phosphate buffer to each bag and prehybridize at 65°C for 1–2 h.
2. Denature an appropriate amount of labeled probe by boiling for 3 min and add to 2 mL of phosphate buffer.
3. Cut open one side of the bag and add 2 mL of phosphate buffer containing the probe to give a final concentration of 2×10^6 cpm/mL buffer. Reseal the bag, taking care not to trap air bubbles. Spread the probe evenly using a roller, and incubate overnight at 65°C.
4. Rinse the filters in 2× SSC and then wash in 2× SSC preheated to 65°C for 30 min × 2, with gentle shaking.
5. The final wash is 0.2× SSC, 0.1% SDS at 65°C for 15–30 min, depending on the "background" level of radioactivity, which should be ~1–2 cpm.
6. Rinse in 2× SSC and blot between Whatman 3MM paper. Then mount the filters between cling film while still moist. It is important not to let the filters dry out completely if they are to be reprobed, since it can be difficult to elute previously bound probe.
7. Mount the filter and autoradiograph between two intensifying screens at –70°C.
8. Develop the film after an overnight exposure. The locus-specific HVR probes are extremely sensitive and often a 2–3 h exposure is adequate.

4. Applications

4.1. Determination of Correct Paternity

DNA fingerprints hold a significant advantage over standard blood group and protein markers in determination of correct paternity: the individual specificity allows both exclusion and positive confirmation of paternity with a very high degree of certainty. In genetic linkage studies, this is useful in verifying the reported pedigree structure. The analysis is based on the Mendelian pattern of inheritance of fingerprint alleles: apart from relatively rare mutant alleles, each band in an individual's fingerprint can be traced to the fingerprint of that individual's father or mother (Fig. 2). Exclusion of paternity should be confirmed by repeating the fingerprinting using a second restriction enzyme. The power of DNA fingerprinting in paternity testing is greatly increased if several different probes are used sequentially (*see* Note 1). Although a single fingerprint is usually sufficient to confirm or exclude paternity, if two or more fingerprints are examined, further conclusions may be drawn, e.g., possible consanguinity (Fig. 3).

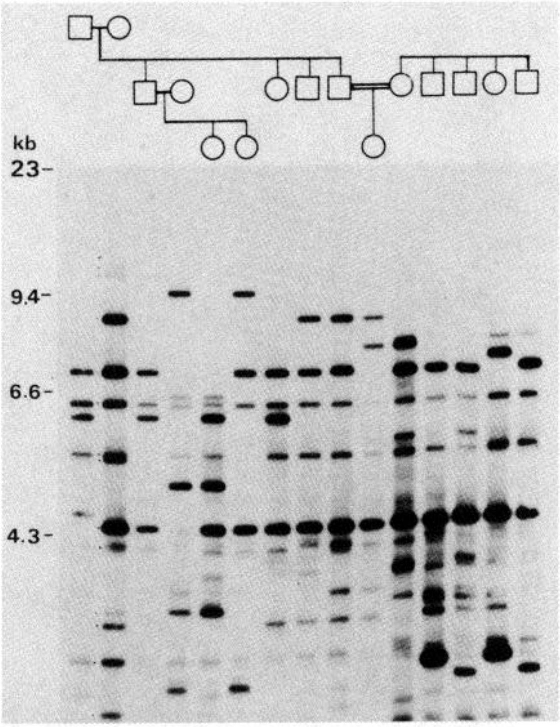


Fig. 2. Mendelian inheritance of DNA fingerprint alleles. DNA fingerprints of various members of a pedigree produced after *Alu* 1 digestion and hybridization to the 3' α HVR probe under conditions of relaxed stringency. The double line (=) between two of the individuals indicates that they are related to each other. Note that the fingerprint bands are inherited in a Mendelian fashion and that each band in an individual can be traced to the fingerprint of either the father or mother.

4.2. Determination of Twin Zygosity

A special case in which DNA fingerprinting has proved useful in estimating degree of relatedness is in the determination of twin zygosity. In pairs of the same sex, it is often not clear if twins are mono- or dizygotic. The question can be settled quickly and unambiguously by DNA fingerprinting: monozygotic twins always share identical fingerprints, whereas the probability that dizygotic twins share identical fingerprints equivalent to that of sib-pairs is in the order of 10^{-7}, using one minisatellite probe.

A pool of locus-specific HVR probes, e.g., six can be used to produce a "reconstituted" DNA fingerprint (Fig. 4). The multilocus Southern blot pat-

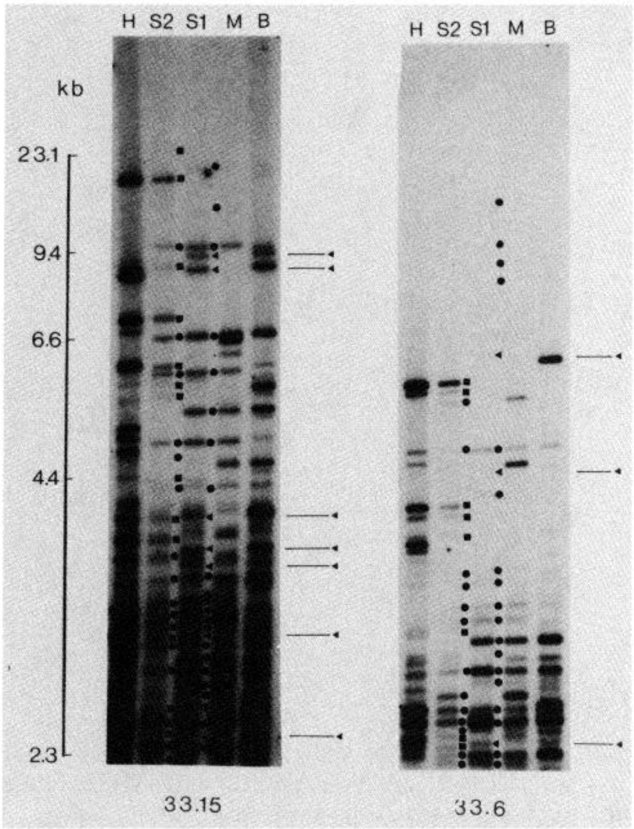


Fig. 3. Prediction of consanguinity using DNA fingerprinting analysis. DNA fingerprints of a family produced by hybridization of *Hinf* 1-digested DNA to the 33.15 and 33.6 probes. Lanes: H = husband of M, S2 = brother of S1, S1 = proband, M = mother of S1, and B = brother of M. ● = Maternal bands that could be resolved in S1 and S2. ■ = Nonmaternal bands that could be resolved in S2. ▲ = Nonmaternal bands that could be resolved in S1 and that were also present in B. DNA fingerprints of S1 and S2 show 49 and 50 resolvable bands, respectively. S1 shares 39/49 bands, whereas S2 shares 27/50 bands with M, giving odds of 9.9×10^{14} to one and 4.5×10^{5} to one in favor of a first degree relationship with M, respectively. All the 23 nonmaternal bands in S2 could be traced to H; S2 is almost certainly the child of H and M. However, of the 49 bands in S1, 10 cannot be traced to either H or M; thus, the reported family relationship is false. Furthermore, the observed number of band-sharing between S1 and M is significantly higher than expected, even for a first degree relationship, which suggests that the father of S1 is related to M. Examination of the DNA fingerprints showed that all the nonmaternal bands in S1's fingerprints were present in M's brother, B. After completion of the study, M volunteered that her brother, B, was the true father of S1.

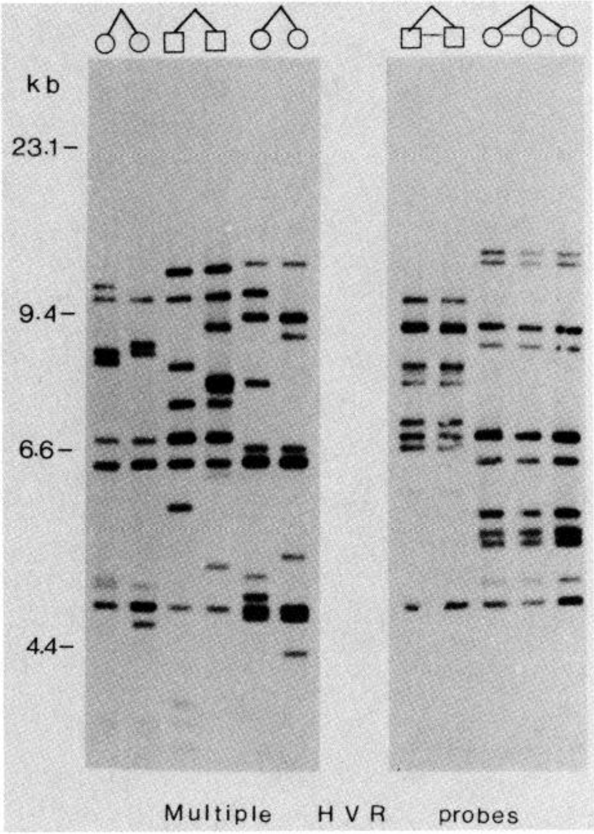


Fig. 4. Assessment of twin zygosity using HVR probes. "Reconstituted" DNA fingerprints generated by hybridization of *Hinf* 1-digested DNA of sets of dizygotic twins (panel on left) and monozygotic twins and triplets (panel on right) with six locus-specific HVR probes (λMS.1, λMS.8, λMS31, λMS43, λMS51, and pλg3).

terns are highly individual-specific; in this case, the cumulative probability of two sibs being identical for all six probes is 0.0004, compared with ~10^{-7} for a DNA fingerprint using one "polycore" probe.

4.3. Tumor Studies

HVR probes have proven to be useful tools in DNA studies of neoplasms and can be applied in two areas: (1) in the assessment of clonality by DNA fingerprinting using the multilocus minisatellite probe, and (2) in the detection of allelic loss using a panel of single-locus probes.

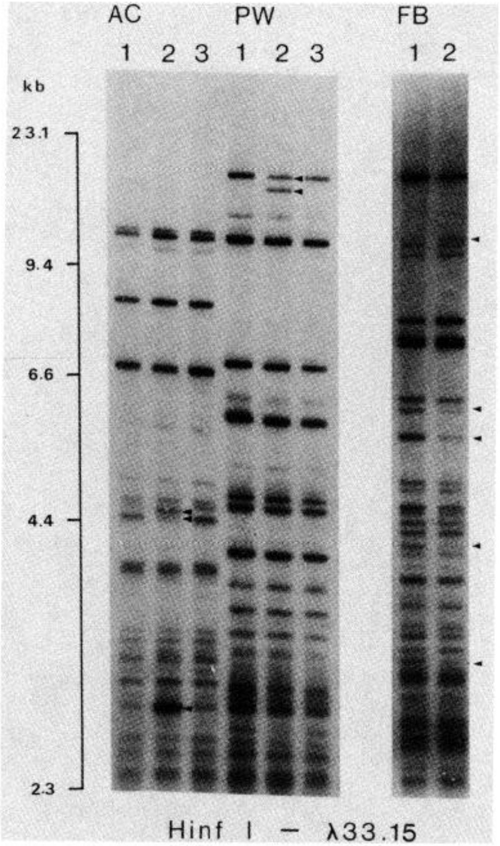


Fig. 5. Assessment of clonality by DNA fingerprinting analysis. Comparison of the constitutional and tumor DNA fingerprints produced after *Hinf* 1 digestion in three patients with colonic cancer. Lanes numbered 1 represent constitutional DNA from peripheral blood leukocytes; 2, tumor DNA; and 3, constitutional DNA from adjacent normal mucosa tissue. The symbol ◀ indicates the position of novel bands and shifts in band intensities seen in the tumor DNA fingerprints.

DNA fingerprinting provides a means by which a large number of dispersed HVR loci can be screened simultaneously. The added combination of a particularly high mutation rate at these loci makes the technique an attractive approach for the detection of somatic mutations in the DNA of tumor cells. The mutations present as either novel fingerprint bands, loss of bands, or both (Fig. 5). Although somatic mutations are readily detectable by this technique, it is not possible to localize any detectable change to a particular part of the genome. However, somatic mutations giving rise to novel bands

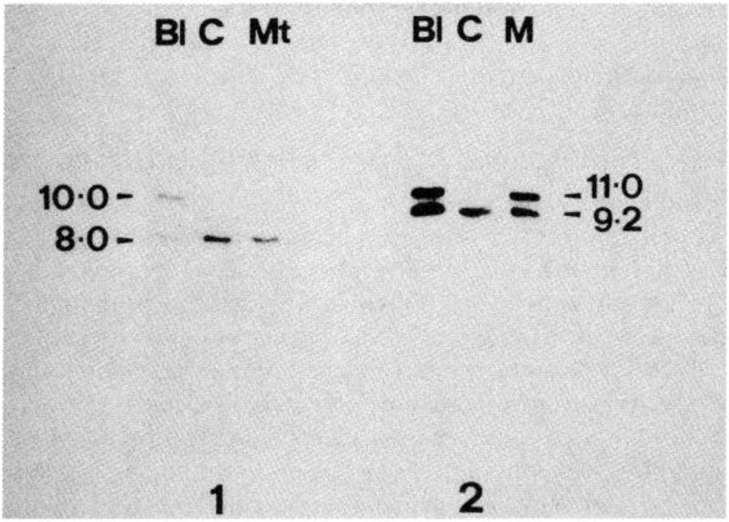


Fig. 6. Detection of allele loss in two cases of stomach cancer. Lanes labeled BI = constitutional DNA from peripheral blood leukocytes; C = DNA from tumor tissue; Mt = DNA from metastatic lymph node, and M = DNA from adjacent normal gastric mucosa. In case 1, DNA was digested with *Eco*RI and hybridized to pMUC10, which detects an HVR on chromosome 1q, and in case 2, *Hinf* 1-digested DNA was hybridized to λMS43, which detects an HVR on chromosome 12q. Comparison of the hybridization patterns shows loss of heterozygosity in both tumor and metastatic tissue DNA in 1, and in tumor DNA in 2.

could be used as clonal markers for tumor cell populations, and should prove useful for the study of tumor progression.

Somatic mutations resulting in deletion of "anti-oncogenes" have been implicated in the pathogenesis of several neoplastic diseases. Owing to their extreme heterozygosity, single-locus HVR probes in the implicated regions are particularly useful for the detection of such deletions, which are identified as a loss of heterozygosity when compared to the DNA from constitutional tissue (Fig. 6).

It is important to recognize that tumor tissues may be comprised of a mixed population of parental and mutant cells, and are often surrounded by "normal" inflammatory cells. Thus, any somatic mutation detected would be more obvious if DNA from tumor tissue is electrophoresed in a track adjacent to that of an equivalent amount of constitutional DNA.

A potential of source error lies in possible tumor-specific methylation of DNA. Restriction endonucleases, such as *Hinf*I and *Sau* 3A1, which are commonly used in DNA fingerprinting, are sensitive to CpG methylation at their

recognition sites, which could result in differential restriction digestion patterns of otherwise identical stretches of DNA. This problem can be obviated by choosing restriction enzymes that are not sensitive to methylation, e.g., *Alu* I and *Hae* III.

4.4. Unambiguous Identification of Individuals

The most celebrated application of DNA fingerprinting technology has been in the field of forensics, where its use has resulted in the perpetrators of a number of violent crimes being brought to book. This application is based on the individual specificity of DNA fingerprints: the probability that two unrelated individuals share identical fingerprint patterns being infinitesimal. Furthermore, because of the repetitive structure of minisatellites, fingerprinting probes are very sensitive; sufficient DNA can be recovered from a drop of blood, or 15 hair root follicles, or 5 µL of semen, to generate full DNA fingerprints. Thus, tiny traces of tissue left by a criminal at the scene of the crime allow irrefutable evidence of his/her identity to be obtained.

The principle underpinning this forensic application has been extended to medical science, particularly in the identification of posttransplant cell populations. DNA fingerprint analysis of marrow tissue or peripheral blood leukocytes after transplantation allows assessment of engraftment; a successful graft retains the fingerprint of the donor, whereas reversion to pretransplant fingerprint indicates a failure of the graft to "take" (Fig. 7).

4.5. Genetic Linkage Analysis

The application of restriction fragment length polymorphisms (RFLPs) to linkage analysis has become commonplace. There are, however, difficulties inherent in this approach. Conventional RFLPs (single-base substitutions or insertion-deletion) are merely dimorphisms, and therefore can have no more than 50% heterozygosity; this limits their informativeness in linkage studies. Minisatellites, with their extremely high heterozygosity, make much more informative genetic markers. DNA fingerprinting using polycore probes would allow many such markers to be studied at once, and in large families, one can screen for hypervariable fragments that apparently cosegregate with an inherited disease. However, several constraints limit the utility of fingerprinting for this application.

Since there is no way of determining which fingerprint bands in one individual represent the locus of a band in an unrelated individual, the usual practice of combining linkage data from a number of small pedigrees is impossible. Thus, in order for the analysis to be informative, the number of meioses studied in a single pedigree must be sufficiently large to yield results of statistical significance (at least 10 phase-known meioses); such large horizontal pedigrees are not common. However, once such a "linked" fragment

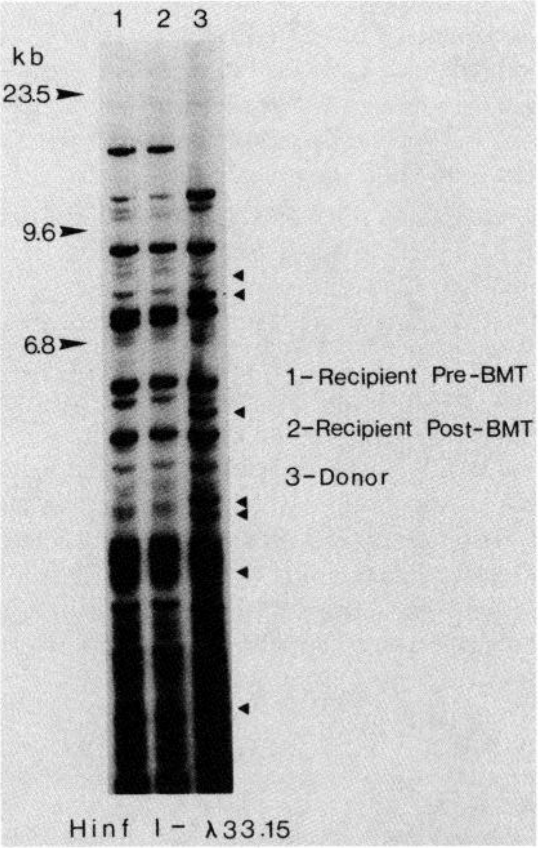


Fig. 7. Assessment of bone marrow engraftment by DNA fingerprinting analysis. DNA was digested with *Hinf* 1 and hybridized with probe 33.15. Lane 1 = recipient prebone-marrow transplant (BMT); 2 = recipient post-BMT; 3 = Donor (brother of recipient). Twenty-five DNA fragments could be resolved in the patient, and 21 in the donor, of which 14 were shared between the siblings and seven were specific to the donor. No donor-specific bands were present in the post-BMT DNA fingerprint, indicating that the engraftment is autologous.

is identified, it can be cloned to provide a locus-specific hypervariable probe suitable for linkage analysis in other affected families.

Alternatively, a panel of highly informative single-locus probes can be generated from a polycore probe. Such markers are now in widespread use in human genetics, and have been used to localize the genes for adult polycystic kidney disease and inherited atopic illness.

5. Notes

1. Maximum information can be derived from a filter, if it is sequentially hybridized with all three "polycore" minisatellite probes, with elution of probe between hybridizations. Provided the membrane is not allowed to dry out completely, probe previously bound to a nylon membrane is easily removed by the following method.
   a. Incubation in 0.4*M* NaOH, 45°C for 30 min followed by neutralization in 0.2*M* Tris-HCl, pH 7.5, 0.1% SDS, 0.1× SSC at 45°C for 30 min.
   b. After a brief rinse in distilled water, the filters are ready to be rehybridized.
2. Although DNA fingerprinting is a robust technique, there are a number of potential sources of error. The most prominent of these is incomplete digestion of DNA samples, leading to the appearance of spurious fingerprint alleles. As for any Southern blotting of genomic DNA, it is a sensible precaution to use a fivefold excess of restriction enzyme and to allow the digests to incubate for 6–12 h. Any DNA fingerprint with a noticeably larger than average number of bands might have resulted from partial digestion, and should be viewed with suspicion.

References

1 Jeffreys, A J., Wilson, V., Wong, Z., Royle, N., Patel, I , Kelly, R., and Clarkson, R. (1989) Highly variable minisatellites and DNA fingerprints *Biochem Soc Symp* **53**, 165–180.

2. Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985) Hypervariable "minisatellite" regions in human DNA. *Nature* **314**, 67–73

3 Nakamura, Y., Leppert, M , O'Connell, P , Wolff, R , Holm, T , Culver, M., Martin, C., Fujimoto, E., Hoff, M , Kumlin, E , and White, R. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**, 1616–1632.

4 Jeffreys, A. J., Wilson, V , and Thein, S L (1985) Individual specific "fingerprints" of human DNA. *Nature* **316**, 76–79

5. Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A. S., and Christophe, D (1987) A sequence in M13 phage detects hypervariable minisatellites in human and animal DNA. *Science* **235**, 683,684.

6. Jarman, A, Nicholls, R. D , Weatherall, D J., Clegg, J B , and Higgs, D. R. (1986) Molecular characterization of a hypervariable region downstream of the human α-globin gene cluster. *EMBO J* **5**, 1857–1863

7 Wong, Z , Wilson, V , Jeffreys, A. J , and Thein, S L (1986) Cloning a selected fragment from a human DNA "fingerprint": isolation of an extremely polymorphic minisatellite. *Nucleic Acid Res.* **14**, 4605–4616.

8 Wong, Z., Wilson, V., Patel, I., Povey, S., and Jeffreys, A. J. (1989) Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* **51**, 269–288

CHAPTER 23

DNA Fingerprinting and Forensic Medicine

Karen M. Sullivan

1. Introduction

DNA fingerprinting without doubt represents one of the most significant advances in forensic science this century. Central to this technology, which is based on the analysis of the genetic component of cells, is the use of DNA probes to regions of the human genome that exhibit great variability between individuals *(1)*. These probes fall into two main categories. The first group comprises those that can detect a large number of these "hypervariable" loci simultaneously, namely multilocus probes (MLPs). On autoradiography, these give rise to a band pattern that is reminiscent of the bar code on supermarket goods, the main advantage of which is that a single such test provides a lot of information very rapidly. MLPs are, therefore, the probes of choice when the amount of material for testing is not limiting, e.g., a blood sample for paternity testing. In many forensic cases, however, the material evidence available for testing is minute, such as a few hair roots or a tiny semen stain, and the situation is often complicated by the presence of tissue from more than one person. In such cases, probes that detect only a single region in the human genome are used, i.e., single-locus probes (SLPs) *(2)*. Such probes have an advantage over MLPs in that they are very much more sensitive, needing a much smaller quantity of DNA to provide a result. In addition, the limited number of bands they detect in a sample makes it possible to resolve individual contributions to a DNA fingerprint obtained from a mixture of components. The main drawback to the use of SLPs is that each test yields

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

only a limited amount of information, so several different SLPs must be used consecutively to generate a high degree of certainty of a match of two samples, thus protracting the time-scale of the analysis.

Not only are the loci detected by MLPs and SLPs very variable, but also they are inherited in a Mendelian fashion, so all the bands in a child's DNA fingerprint are inherited from his or her parents. Hence, the more related two people are, the greater is the number of bands shared in their DNA fingerprints. This has led to the establishment of DNA fingerprinting as the definitive method of relationship testing in both civil and criminal paternity disputes, and also in cases in which immigration rights are claimed on the basis of family relationships *(3)*.

The use of DNA fingerprinting in forensic medicine will be discussed in this chapter. Details of the fingerprinting technique can be found in Chapter 22, this volume.

2. Materials

1. Apparatus required for the DNA extraction processes comprises disposable microcentrifuge tubes, 0.2- and 1.0-mL pipet tips, Petri dishes, Universal tubes, scalpels, and swabs. In addition, access to a microcentrifuge and a vacuum line will be necessary, and it is advisable to carry out the initial stages of extraction, i.e., up to the ethanol precipitation step, in a Class II containment unit.
2. Extraction buffer (×2): 20 m*M* Tris-HCl, pH 8.0, 20 m*M* EDTA, 0.2*M* NaCl, 4% SDS (sodium dodecyl sulfate).
3. Solvents: 100% ethanol, 80% ethanol, phenol/chloroform/isoamyl alcohol (25/24/1), chloroform/isoamyl alcohol (24/1), sterile distilled water.
4. Stock solutions (made up as specified in the text, as required): 1*M* Tris-HCl, pH 8.0; 1*M* DTT (dithiothreitol); 0.5*M* EDTA, pH 8.0; 1*M* NaCl; and 1*M* trisodium citrate.
5. Saline sodium citrate (20× SSC): 3*M* NaCl, 0.3*M* trisodium citrate, pH 7.0.
6. Glycogen, 20 mg/mL.
7. Proteinase K, 10 mg/mL.
8. Restriction enzyme buffer made up to suppliers' specifications.

3. Methods

3.1. DNA Extraction from Forensic Samples

Techniques for the extraction of DNA from a variety of forensic samples are described below. Following extraction, and before proceeding with the

DNA-fingerprinting analysis, it is important to assess accurately the quantity and quality of the DNA recovered. The quantity is best determined by removing a small aliquot for fluorimetry. The quality, i.e., the extent to which the DNA has degraded, can be assessed by removing two small aliquots of the DNA—one before and one after restriction enzyme digestion—and running them on a 0.7% mini agarose gel against standard DNA. This allows a visual estimation of the ratio of high-mol-wt DNA to degraded DNA, and also provides a check on whether the digestion has gone to completion (*see* Note 1). Loadings for the analytical gel should be adjusted appropriately, so all samples contain approximately the same amount of high-mol-wt DNA.

3.1.1. DNA Extraction from Whole Blood Samples

1. A 1.0-mL aliquot of whole blood yields sufficient DNA for several MLP analyzes. Freeze the blood at –20°C until required.
2. Thaw, make the volume up to 1.5 mL with 1× SSC, mix, and pellet the white cells by a short spin in a microcentrifuge.
3. Remove the supernatant (approx 1.0 mL) and repeat step 2.
4. Resuspend and incubate the white pellet in 0.4 mL of 10 m*M* Tris-HCl, pH 8.0, 10 m*M* EDTA, 100 m*M* NaCl containing 2% SDS, 20 µg/mL proteinase K, and 39 m*M* DTT, for 3 h at 37°C.
5. Purify the DNA by two phenol/chloroform extractions.
6. Precipitate the DNA using ethanol: Add 0.1 vol of 2*M* sodium acetate and 2.5 vol of absolute ethanol. DNA should then spool out and become clearly visible as a white, diffuse pellet.
7. After a short spin in a microcentrifuge, remove the supernatant and resuspend the pellet in 0.2*M* sodium acetate. Add ethanol and then reprecipitate.
8. Wash the pellet in 80% ethanol, spin to repellet, and remove as much of the supernatant as possible. Vacuum-dry the pellet.
9. Resolvate the pellet in HinfI restriction enzyme buffer. This is best done at 37°C for 1–2 h, with intermittent gentle vortexing. It is vital to ensure complete resolvation before proceeding to digestion.
10. To minimize the risk of obtaining partial digestions, incubate overnight if possible, with a large number of enzyme units (up to 40 U/mL of whole blood) (*see* Note 2).

3.1.2. DNA Extraction from Blood Clots

1. Cut approx 0.4 mL of blood clot into small pieces on a sterile Petri dish; then transfer to a sterile microcentrifuge tube.
2. Wash the clot in 1 mL of 1× SSC; then remove and discard the supernatant.

3. Resuspend the clot in 0.4 mL of 0.2*M* sodium acetate and add 20 µL of 10 mg/mL proteinase K, 20 µL of 1*M* DTT, and 30 µL of 10% SDS.
4. Incubate at 56°C for 6–24 h.
5. Phenol-extract twice and chloroform-extract once.
6. Add 1 µL of 20 mg/mL glycogen prior to the first ethanol precipitation, and then proceed to precipitate, wash, dry, and cut the DNA as described for whole blood. If the DNA does not spool out and become visible on the addition of the ethanol, leave at –20°C for 1 h before centrifugation. This additional step is required for the majority of forensic samples.

3.1.3. DNA Extraction from Muscle and Fetal Tissue and CVS Specimens

1. Cut approx 50–100 µL of tissue from the most central part of the available muscle biopsy. Chop into small pieces on a sterile Petri dish; then transfer to a microcentrifuge tube for processing.
2. Cut approx 30–50 µL of fetal limb-muscle tissue into small pieces. Cut from frozen fetal material, rather than letting the tissue thaw first; transfer to a microcentrifuge tube; and commence processing immediately. (Fetal tissue tends to degrade very rapidly after freeze-thawing.)
3. Pellet approx 100 µL of solid material from a CVS (chorion villus sample) specimen into a microcentrifuge tube, and remove the supernatant. If some of the pieces are a bit large, mince in the tube, using the fine end of a Pasteur pipet.

DNA is subsequently extracted from all of these tissues as described for blood clots.

3.1.4. DNA Extraction from Blood Stains: Direct Lysis

1. Cut the stain (approx 1 cm^2) into small pieces on a sterile surface, and then transfer to a sterile Universal tube.
2. Lyse in 0.5 mL of 2× extraction buffer, 0.4 mL of water, 40 µL of 1*M* DTT, and 20 µL of 20 mg/mL proteinase K. Incubate at 37°C overnight.
3. Recover as much supernatant as possible, then wash the material with a further 0.2 mL of sterile water and add this to the first supernatant in a microcentrifuge tube.
4. Purify and precipitate the DNA as described for blood clots.

3.1.5. DNA Extraction from Semen Stains and Vaginal, Anal, and Oral Swabs: Differential Lysis

1. Cut the stain or swab head into small pieces and place in a sterile Universal bottle.
2. Lyse in 5.6 mL of 100 m*M* NaCl/10 m*M* EDTA, 0.38 mL of 10% SDS,

and 0.15 mL of 10 mg/mL proteinase K for at least 2 h at 56°C. If there is heavy contamination with epithelial cells, this time should be increased.

3. Remove the supernatant from the material to a microcentrifuge tube, and pellet the sperm heads by spinning for 4 min in a microcentrifuge.
4. Wash the material or swab twice with 1.5-mL aliquots of NaCl/EDTA and pellet the additional sperm heads with the main sample.
5. Second lysis: Resuspend the sperm pellet in 0.4 mL of 2*M* sodium acetate. Add 20 µL of 1*M* DTT, 20 µL of 10 mg/mL proteinase K, and 30 µL of 10% SDS, and incubate at 37°C overnight.
6. Purify and precipitate the DNA as described for blood clots.

3.1.6. DNA Extraction from Hair Roots

1. Remove the roots from the hair shaft and place in a 0.4-mL microcentrifuge tube.
2. Add 100 µL of extraction buffer, ensuring that all hair roots are submerged in the buffer at the bottom of the tube.
3. Incubate at 37°C overnight.
4. Purify and precipitate the DNA as described for blood clots.

3.2. Identification of Matching DNA Fingerprints and Assessment of the Probability of Random Matching

When two DNA fingerprints are deemed to match, either by visual inspection or by accurate size analysis of the bands that make up the fingerprint, there are two possibilities that must be considered:

1. The profiles match because they are from the same individual—in this case, the probability of the match is 1.
2. Alternatively, the profiles may be derived from two different people, and they just happen to match by chance—in this case, the probability is determined by the frequency at which the DNA fingerprint in question may be expected to occur within the population as a whole.

The method of calculation used to evaluate the probability of chance matching (i.e., the second possibility above), is dependent on the type of probe used (SLP or MLP) and, in the case of SLPs, is also dependent on the ethnic origin of the individual being tested.

3.2.1. Forensic MLP Analysis

MLP analyses are most simply and accurately performed as a side-by-side analysis of samples to be compared. In this case, a simple visual inspection reveals the presence of matching profiles. The matching DNA fingerprints are then analyzed in more detail, and all bands larger than 4 kb in the pro-

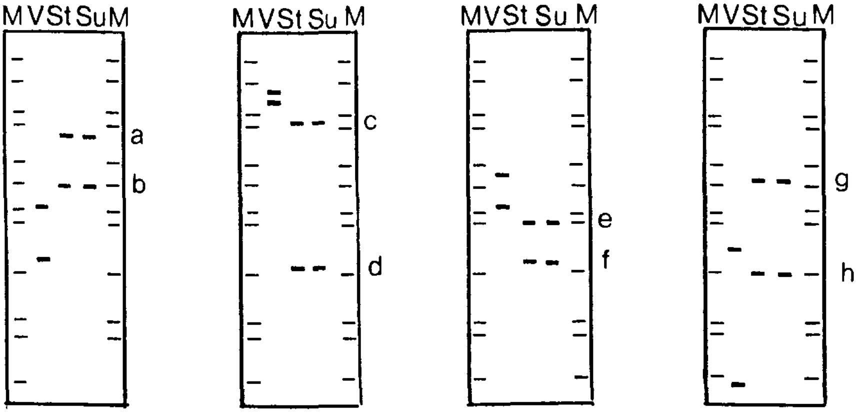
files are scored (this is an arbitrary "cutoff" point below which the bands become more compacted and more difficult to score accurately). All such bands must have a match in both position and intensity on each profile for the two DNA fingerprints to be confirmed as a match. An exception to this rule is when the test sample is partially degraded with respect to the reference sample. In this case, the top few bands of the DNA profile from the test sample may be missing, since high-mol-wt bands are the first to degrade. Such degradation should have been detected at the quality control stage and, with MLP analysis, degraded DNA generally gives a high background to the profile. Taking into account these factors, and the extent of matching at the lower end of the DNA profile, an informed judgment has to be made as to whether or not the DNA fingerprints are matching.

It is particularly important, when performing an MLP analysis, that all samples are completely digested, since partial digests can have a greater number of bands than their fully digested counterparts, and this could lead to erroneous exclusion of matches.

It has been established that, on average, unrelated people share 25% of the scored bands in their DNA fingerprints, i.e., the probability of any one band finding a match in a DNA fingerprint from an unrelated person is 1 in 4. The probability of two bands matching by chance is, therefore, $0.25 \times 0.25 = 0.0625$, or 1 in 16; that of three bands matching by chance is $0.25 \times 0.25 \times 0.25 = 0.015625$, or 1 in 64; and so on. If, for example, 10 matching bands were scored in two DNA profiles from a semen stain and a reference blood sample, the probability of a random match would be $0.25^{10} = 9.536 \times 10^{-7}$. This could be expressed as one chance in 1,048,575 that the profiles match by chance rather than because they are from the same person.

3.2.2. Relationship Testing

The method of calculation is the same as for the forensic MLP analysis, in that band-share values are used to calculate the probability of random matching, but the scoring of bands is different. In forensic cases, two full profiles must be matching, but in paternity testing, the test is to establish whether the alleged father could have contributed all the paternal bands to the child's DNA profile. To do this, the child's DNA profile is first compared to that of his or her mother, and all the bands that find a match in the maternal profile are discounted, leaving to be considered only those bands that must have been inherited from the father. To confirm paternity, all the paternal-specific bands present in the DNA profile of the child must be matched by bands present in the DNA profile of the alleged father. The possibility that the putative father simply matches the paternal bands in the child's DNA by chance is then calculated by raising 0.25 to the power of the number of matching bands.



MS1	MS31	MS43	G3	PROBE
f(a) = 0.04	f(c) = 0.30	f(e) = 0.26	f(g) = 0.16	
f(b) = 0.12	f(d) = 0.11	f(f) = 0.07	f(h) = 0.04	

Fig. 1. Typical forensic SLP-analysis data, showing representative frequencies of allele occurrence (to 2 SD): M, marker track; V, victim's DNA sample; St, DNA extracted from semen stain on exhibit; Su, suspect's DNA sample; f(*x*), frequency of occurrence of allele *x*.

3.2.3. Forensic SLP Analysis

In this kind of analysis, a maximum of eight bands are normally examined per profile, so it is reasonable to size each band accurately by including a standard mol-wt ladder on each test, and to cross-compare the size of each band between profiles. The criteria of a match, if the bands are not perfectly aligned, depends largely on the discretion of the analyst, and may be either a fixed "window" or two or three standard deviations, accordingly.

Once the band sizes are calculated and two DNA profiles are deemed to match, the probability of a chance match is calculated by reference to a data base of allele sizes for each probe and ethnic origin of the person being tested. Using the data base, the frequency of occurrence of each band within the population is established. A schematic representation of data from a forensic SLP analysis is shown in Fig. 1, together with a typical set of allele frequencies. It is clearly seen that the frequency of occurrence of each allele is relatively high, with some bands occurring in almost a third of the population. The key to the generation of the impressive statistics associated with DNA fingerprinting is that each individual frequency is unrelated to the others, i.e., there is no demonstrated linkage between the alleles, and the probability of matching all these alleles by chance is the product, therefore, of

all the individual frequencies. In the case illustrated in Fig. 1, the calculation would be as follows:

$$2(0.04 \times 0.12) \times 2(0.30 \times 0.11) \times 2(0.26 \times 0.07) \times 2(0.16 \times 0.04) = 2.59206 \times 10^{-7} = 1 \text{ in } 3{,}387{,}454$$

i.e., the probability of a chance match of the DNA profiles detected is less than 1 in 3.3 million. The factor of 2 is introduced for each pair of alleles because the numbers shown represent allele frequencies and, since we are diploid, we have two chances of inheriting a given allele.

It is necessary to generate a data base for each ethnic group, since the distribution of the alleles detected by any given probe may vary significantly between peoples of different ethnic groups.

4. Applications

4.1. Relationship Testing

Relationship testing by DNA fingerprinting is widely used to resolve, in addition to civil paternity and inheritance disputes, immigration cases, in which proof of biological kinship to a resident of the UK often entitles the applicant to enter and reside in Great Britain. Increasingly, however, the technology is being used to provide evidence in cases of criminal paternity *(4)*.

There are several scenarios in which DNA fingerprinting can generate data that are of considerable evidential value. In rape cases in which no immediate evidence of intercourse, such as vaginal swabs or semen stains, is available, but in which the victim subsequently conceives, paternity testing of the offspring can at least demonstrate that intercourse with the suspect occurred (in cases of positive paternity). In cases in which an abortion has been performed, the fetus can be used as the source of DNA for testing (*see* Note 3), providing the method of abortion allows fetal material to be distinguished from maternal tissue. When the mother has carried to term and given birth to a child, EDTA or clotted blood samples are generally provided from the mother, child, and alleged father (*see* Note 4), although if the child is newborn, a blood stain from a heel prick is sometimes submitted in place of a blood sample, to minimize trauma to the child. In all cases, sufficient DNA (approx 1–2 µg) for an MLP analysis to be performed is usually available. For relationship testing, it is essential that the family group is tested side-by-side for ease of analysis.

The most outstanding progress made by the technology in this general area is in providing evidence of paternity in incest cases. Because of the high degree of relatedness of offspring from incestuous relationships, conventional serological and biochemical methods have, on the whole, been unable to

provide strong evidence of paternity in such cases. However, using MLPs, or, preferably, a serial combination of two different MLPs, sufficient information can be gained to provide impressive statistics in favor of paternity, even when the alleged father of the offspring is a first-degree relative of the mother. The only "loophole" in this approach is that, when one exists, a brother of the accused is often named as an alternative father of the child. Because of the high proportion of shared bands between brothers (an average band share of 62.5%), it is better to test the brother, if possible, than to calculate the relative statistics of the likelihood of his being the child's father.

When a termination of pregnancy is requested on the grounds that a child may have been conceived as a result of rape or incest, it is possible to determine the paternity of the unborn child by performing DNA fingerprinting on DNA extracted from a CVS. The results of such tests can be gained in a suitable time scale to allow abortion to proceed should an unfavorable paternity be established.

Relationship testing "in reverse" may be used to identify offspring rather than parent, when corpses cannot be identified by conventional means. When identification is impossible because of decay of the deceased, lack of dental records, or the circumstances under which the individual died (e.g., explosions, crashes, or industrial accidents), identity can be established when putative parents or close relatives are available for testing. DNA from the body under investigation is best extracted from a deep biopsy from the thigh muscle, i.e., not from subcutaneous muscle (*see* Note 5).

A further application of relationship testing using DNA fingerprinting exploits the exception to the rule that every individual has a different genetic makeup: namely, identical or monozygotic twins. Identical twins possess identical DNA fingerprints, which provides a definitive method of establishing zygosity when one twin requires an organ or bone-marrow transplant *(5)*.

4.2. Identification of Assailants in Sexual Crime

Many workers in the field of forensic science would argue that it is within this area that DNA-fingerprinting technology has made its most significant contribution *(6)*. The forensic samples in such cases most often consists of vaginal, anal, or oral swabs, and semen stains (*see* Note 6). The nature of the swab samples, and the way in which they are taken, results in heavy contamination of the semen with epithelial cells from the lining of the vagina, anus, or mouth. This is also true, but to a lesser extent, with semen stains made following penetration. The result is that the forensic scientist is faced with analyzing a mixture of material from both the victim and the assailant. In some cases, this is complicated further by the presence of semen from more than one assailant or from voluntary intercourse with another partner prior

to the assault. Using conventional grouping techniques, this is a very difficult, or sometimes impossible, task. However, using SLPs, a maximum of two bands are detected per person, per probe—there are two bands per SLP, because, although the probe detects only a single locus, there are two copies of that region present: one inherited from the mother and one from the father. This means that, given reference DNA samples from all parties involved, it is possible to assign their individual contributions to the DNA profile obtained from the mixture, and hence to identify or eliminate the suspects (*see* Note 7).

In most cases, the picture can be simplified somewhat by elimination of the contaminating epithelial cells prior to processing the samples, using a "differential lysis" procedure. Sperm heads are resistant to lysis in SDS in the absence of DTT, allowing epithelial cells to be lysed while sperm heads remain intact. Sperm heads can then be pelleted by centrifugation and separated from the epithelial DNA in solution prior to lysis in the presence of DTT. In cases of particularly heavy epithelial-cell contamination, there is sometimes residual DNA from the victim, for which reason a reference blood sample from the victim is generally requested. This allows the victim's contribution to the DNA profile from the exhibit to be excluded from consideration. In some cases, however, there is a specific requirement that the victim's bands be present in the DNA profile from the exhibit, since this eliminates any possibility of challenging the evidence on the basis of switching or misidentification of the exhibit. If this is required, a direct lysis of both cellular components together is performed, as for blood stains.

It is often forgotten that this technology not only provides an excellent means of identification, but it also provides an even more rapid, and equally conclusive, method of exclusion. This is of particular use in extended police investigations, specifically rape/murder cases, in which a very large number of suspects are being screened. A preliminary screen by blood grouping makes a considerable reduction in the number of suspects to be screened in the first instance. This is advisable, since serology is less expensive and quicker than DNA fingerprinting. Screening the remaining suspects, of whom there may still be a considerable number, would be a daunting task by conventional police work, whereas DNA testing can lead to very rapid, and comparatively inexpensive, elimination of a large number or all of the suspects.

4.3. Identification of Assailants in Violent Crimes

In many incidents of violent crime (including sexual crime), blood from the victim is spilled on the clothes of the assailant. Such garments, recovered at a later date, may then be used to incriminate the suspect. DNA extracted from blood stains generally remains in a high-mol-wt form for some time after the incident, primarily because stains have a large surface area and dry

rapidly. Moisture is one of the primary agents in the degradation of DNA, and, once a sample becomes badly degraded, it can no longer be used successfully for DNA fingerprinting.

In a number of cases, the assailants are themselves wounded in the struggle, and leave traces of their own blood at the scene of the crime. Dried blood can be recovered from almost any type of surface without detriment to or alteration of the DNA profile (*see* Note 8).

In some instances, the victim pulls out some of the hair from the attacker, in which case DNA extracted from the hair roots can be used for identification. Unfortunately, shed hair, which is often found at the scene of a crime, cannot be used effectively for testing. Such hairs have little or no cellular material attached to the base of the hair shaft, and are not, therefore, a source of DNA (*see* Note 9).

4.4. Accident Investigation

In accident investigations following disasters, it can be of use to investigators to determine the specific location of victims on impact, and the direction in which they were subsequently thrown or fell. In this case, it is possible to match reference samples from survivors or victims to blood stains found at different locations on various items of wreckage.

On a smaller scale, DNA fingerprinting can sometimes be used following car crashes to match the blood stains on glass, interior, or chassis to those of the occupants of the car. Crash investigators can then assess whether the driver was, in fact, the insured party, in the absence of independent witnesses.

5. Notes

1. When a digest has not gone to completion, the most probable cause is the presence of residual contaminants from the forensic sample, e.g., dye molecules, which inhibit the restriction enzyme. In this case, the DNA should be phenol/chloroform-extracted another two times and reprecipitated before attempting to redigest the DNA with a greater number of restriction enzyme units. Stains on very dark or black material are often refractory to restriction enzyme cleavage, and it is worth including additional purification steps in the first isolation procedure.
2. A single-locus probe MS51 (D11S97) can be used to check the extent of digestion, e.g., when a new batch of enzyme is being tested. MS51 hybridizes to a DNA locus that is located on a restriction fragment bounded by a digestion-resistant HinfI site; therefore when this fragment cuts to completion, the loci detected by the test probes are very likely to be fully cut.
3. It is important that fetal material is frozen immediately following abortion and kept frozen at –20°C until it is required for testing, since fetal

tissue degrades very rapidly. Freeze-thawing also accelerates degradation, and should be minimized. If fetal tissue is dispersed through maternal tissue, it is best to take several biopsies from the products of abortion, sampling from the paler areas of tissue.

4. Clotted blood samples usually yield sufficient DNA for MLP analysis, but EDTA samples are preferable, since they give higher yields of DNA, are easier to process, and are less susceptible to bacterial contamination.
5. Other postmortem tissues may be used for DNA fingerprinting, such as hair roots, bone marrow, or blood, but after a few days, the deep muscle samples are the best source of high-mol-wt DNA.
6. It is critical that swabs be air-dried prior to sealing and storage—if stored damp for any length of time, the DNA may be partially or wholly degraded before processing. It is also important to minimize freeze-thawing, so, if the swabs are frozen prior to transport, they should be kept frozen in transit to the testing laboratory. Anal swabs degrade extremely rapidly, because of the high bacteria content of the sample, and should be processed as rapidly as possible.
7. In forensic science, where SLP testing prevails, it is more usual to refer to "DNA profiles" rather than "DNA fingerprints." The terms refer to exactly the same process, it is simply that the term "fingerprint" implies uniqueness, and SLP testing only occasionally generates the kind of statistics that one would equate with uniqueness. For this reason, it seems less misleading to use the term "DNA profiling."
8. If blood or semen stains are found on fabric, the fabric itself can be cut up and the DNA extracted directly. Where the stain is made on a nonporous surface, the biological material can be removed by scraping with a scalpel blade if there is a heavy deposit, or by swabbing the area with a slightly dampened swab if there is a thinner film. The swab material can then be processed by the normal methods. On surfaces such as wood, the stained area can be chipped or splintered off, or swabbed as for metal surfaces. If the stain is on plant or vegetable material, the best method is to soak the biological material off the substrate, and remove it to a clean tube before processing.
9. There is considerable variation between individuals in the quantity of DNA that is yielded from their hair roots. The number of roots required to obtain a DNA profile varies from one to 10 freshly pulled hairs, or even more if they have been stored and may be partially degraded. Head, body, eyebrow, and pubic hairs are all suitable sources of DNA for testing.

References

1 Jeffreys, A. J., Wilson, V, and Thein, S. L. (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* **314,** 67–73

2 Wong, Z, Wilson, V, Jeffreys, A J, and Thein, S. L. (1986) Cloning a selected fragment from a human DNA fingerprint. Isolation of an extremely polymorphic minisatellite. *Nucleic Acids Res* **14,** 4605–4616

3 Jeffreys, A. J, Brookfield, J F Y, and Semeonoff, R (1985) Positive identification of an immigration test-case using human DNA fingerprints. *Nature* **317,** 818,819.

4 Rittner, C., Shacker, U, Rittner, G., and Schneider, P. M. (1988) Application of DNA polymorphisms in paternity testing in Germany: Solution of an incest case using bacteriophage M13 hybridisation with hypervariable minisatellite DNA *J Adv Forensic Haemogenet* **2,** 388–391.

5 Jones, L, Thein, S L, Jeffreys, A. J, Apperley, J. F., Catovsky, D, and Goldman, J M (1987) Identical twin marrow transplantation for 5 patients with chronic myeloid leukemia: Role of DNA fingerprinting to confirm monozygosity in 3 cases *Eur J Haematol.* **39,** 144–147.

6 Gill, P, Jeffreys, A J, and Werrett, D. J (1985) Forensic Applications of DNA 'fingerprints.' *Nature* **318,** 577–579.

CHAPTER 24

The Detection of Point Mutations in Hemoglobin Defects Using Allele-Specific Oligonucleotide Probes

Swee Lay Thein

1. Introduction

The genetic disorders of hemoglobin, notably, sickle cell anemia and the α- and β-thalassemia, are the commonest genetic diseases in humans. Furthermore, the majority of these mutant globin genes, particularly those causing β-thalassemia, are owing to point mutations that do not involve cleavage sites for restriction enzyme, which means that allele-specific oligonucleotide probe hybridization has become indispensable for the direct detection of these point mutations.

These allele-specific oligonucleotides or ASOs refer to synthetic oligonucleotides whose sequences have been designed to be specific to a short stretch of the human genome in the region of the mutation (*see* ref. *1* and Chapter 7). For the detection of such a mutation, a pair of oligonucleotides are synthesized; one of which is completely homologous to the mutant sequence and the other to the normal sequence, so that there is a single base mismatch between the normal oligonucleotide probe and the mutant sequence, and vice versa. Initially, when ASOs were used for hybridization of genomic DNA immobilized in dried gels, they were designed to be between 19 and 22 bp. This is short enough to differentiate between a perfectly matched hybrid and one with a single base mismatch, and yet long enough so that the sequence detected is unique in the human genome. However, it is now pos-

From: *Methods in Molecular Biology, Vol. 9 Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

sible to achieve a very high degree of enrichment of the target sequence by an in vitro amplification of genomic DNA using the polymerase chain reaction (PCR) *(2)*. This has allowed ASOs of shorter lengths to be used. Furthermore, the increased sensitivity has made it possible to dot-blot the amplified target DNA sequence onto a membrane that is then hybridized to ASOs labeled with ^{35}S, or nonradioactive chemicals, instead of ^{32}P.

ASO probe hybridization depends on the appropriate choice of hybridization and washing temperatures, which exploits the difference in thermal stability between perfectly matched hybrids and those with a single base mismatch. This temperature varies with the length of the probe and its GC content. However, the use of tetramethylammonium chloride (Me_4 NCl), which binds selectively to AT bp and eliminates the preferential melting of AT vs GC bp, has made it possible to control the stringency of hybridization solely as a function of the probe length. Therefore, one can now screen a DNA sample with a panel of ASOs using the same hybridization and washing temperatures.

The procedure for 5' end-labeling ASO with [γ-^{32}P]-ATP and hybridization to (i) genomic DNA immobilized in dried gels and (ii) dot blots of amplified genomic DNA is described.

2. Materials

2.1. Apparatus

2.1.1. Electrophoresis Boxes for Running Horizontal Agarose Gel

1. Slab gel dryer, e.g., BioRad, model 1125 B or Hoefer Scientific Instruments, Dry Gel Sr., model SE1160.
2. X-ray film and cassettes are required for autoradiography. Films such as Kodak XAR 5 or Fuji RX are suitable. Cassettes should be fitted with a calcium tungsten intensifying screen, e.g., Dupont Cronex Lightning plus.
3. Apparatus for automated amplification of genomic DNA, e.g., DNA thermal cycler by Perkin-Elmer (Cetus).
4. Dot-blot apparatus, e.g., BRL, hybridot manifold, or Schleicher and Schuell, Minifold 1.
5. Vertical electrophoresis box for running polyacrylamide gel to separate radioactively labeled ASOs, e.g., Model VCV, No. 62000 by International Biotechnologies Inc., New Haven, CT.
6. Sorvall centrifuge, 15-mL Falcon tubes, and 1-mL syringes for separating labeled ASOs by Sephadex G25-50 spun column chromatography.

2.2. Reagents and Solutions

2.2.1. Preparation of DNA Gels

1. Agarose gel electrophoresis. A type I low EEO agarose is satisfactory, e.g., Sigma No. A-6013, and the electrophoresis buffer is Tris-acetate EDTA (TAE). Prepare a 50× stock, using Tris 242.3 g, NaAc•$3H_2O$ 136.1 g, and EDTA 3.72 g/L, and adjust pH to 8.3 with glacial acetic acid.
2. Restriction endonuclease buffers. These are made up as 10× stock solution according to the manufacturer's instructions. Many manufacturers also provide a 10× stock reaction buffer with the enzyme.

2.2.2. Preparation of Allele-Specific Oligonucleotide Probes

These oligonucleotides should be synthesized with a 5' OH-end.

1. Radioactive nucleotide, e.g., [γ-^{32}P]-ATP (>3000 Ci/mmol, Amersham 15068) or [γ-^{35}S]-ATP (Amersham S.J. 318, >600 Ci/mmol).
2. Kinase buffer: This is prepared as a 10× stock that is 670 m*M* Tris-HCl, pH 8, 100 m*M* $MgCl_2$, and 100 m*M* dithiothreitol (DTT).
3. Loading buffer for the kinase reaction: 0.05% Xylene cyanol, 0.05% bromphenol blue, 20 m*M* Tris-HCl, pH 7.5, 1.0 m*M* EDTA, and 8*M* urea.
4. Sephadex G25-50 suspension in 10 m*M* Tris-HCl, pH 8, 1 m*M* EDTA.

2.2.3. Hybridization and Washing

1. Hybridization buffer: 5× SSPE, 0.1% SDS, and 100 µg/mL yeast tRNA. Make up 20× SSPE stock using 174 g NaCl, 27.6 g Na H_2PO_4•H_2O and 7.4 g EDTA/L H_2O. Adjust pH to 7.4 with NaOH.
2. 20× SSC stock solution: 175.3 g NaCl and 88.2 g sodium citrate/L H_2O. Adjust pH to 7 with NaOH.
3. T-MAC wash solution: 3*M* tetramethylammonium chloride [$(CH_3)_4NCl$] (Aldrich T1, 952-6), 50 m*M* Tris-HCl, pH 8, 2 m*M* EDTA, 0.1% SDS.

2.2.4. Polymerase Chain Reaction (PCR)

1. dNTP mix. We use deoxynucleotide triphosphates (dNTPs) from Boehringer Mannheim, dGTP (Cat. No. 104094), dATP (Cat. No. 103977), dTTP (Cat. No. 104264), and dCTP (Cat. No. 104035), made up in deionized sterile H_2O, and adjusted to pH 7.5 using 2*M* KOH. These are prepared as four separate neutralized 10 m*M* solutions and stored at –70°C. The working solution is a 1:10 dilution of a 10× stock (800 µ*M* final concentration of total dNTPs) prepared from the four separate dNTPs and stored at –20°C.

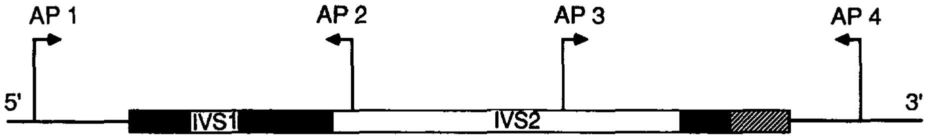


Fig. 1. Representation of the β-globin gene with the position of the primers used in amplification by PCR. PCR primers AP1/AP2 encompass a 916-bp fragment, including the 5' flanking region, exons 1 and 2, and AP3/AP4 encompass a 708-bp fragment, including part of IVS-2, exon 3, and the 3' flanking region of the β-globin gene. The sequences of these primers 5'–3', are:

AP1 – 5'-CGATCTTCAATATGCTTAC-3'
AP2 – 5'-CATTCGTCTGTTTCCCATTCTA-3'
AP3 – 5'-CAATGTATCATGCCTCTTTGCAC-3'
AP4 – 5'-GGCATAGGCATCAGGGCTG-3'

2. Oligonucleotide PCR primers. The majority of the β-thalassemia mutations are concentrated in two regions of the β-globin gene sequence. These are amplified using the two sets of primers AP1 and AP2, AP3 and AP4, as shown in Fig. 1. These are prepared as 2-μ*M* stock solutions in distilled water and stored at –20°C. The working solution is 1:10 dilution to give a final concentration of 0.2 μ*M*.
3. 10× PCR buffer: 500 m*M* KCl, 100 m*M* Tris-HCl, pH 8.4, 15 m*M* $MgCl_2$, 200 μg/mL gelatin.
4. *Taq* DNA polymerase (Amplitaq,® Cetus). We use 2 U/100 μL reaction.

3. Methods

3.1. Restriction Digest of DNA and Immobilization in Dried Gels (see Note 1)

1. Completely digest 10 μg genomic DNA of the patient together with 10 μg genomic DNA of known positive and negative controls with *Bam*HI in a suitable buffer.
2. Electrophorese the digested DNA, including those of positive and negative controls, with λ Hind III marker in a 0.8% agarose gel in TAE buffer overnight.
3. After electrophoresis, immerse the gel in 1 μg/mL ethidium bromide for approx 5 min and then photograph the gel on a UV source. The gel can be trimmed at this stage to remove any unused lanes.
4. Rinse the gel with water and place on two sheets of Whatman 3MM paper. Then transfer the gel with its backing of Whatman paper to a slab gel dryer. Overlay with cling film and cover with neoprene rubber sheet, which is part of the gel dryer.

5. Dry the gel under vacuum initially without heat. When the gel is almost dry, i.e., when the gel feels flat, set the heater to 60°C and continue drying under vacuum at 60°C for 1 h. Release the vacuum; the gel should be a thin film on the Whatman paper, and can be stored indefinitely at room temperature until needed.

3.2. Amplification of Genomic DNA and Preparation of Dot Blots

1. Mix the following:
   a. 10 µL of 10× PCR buffer.
   b. 10 µL of 8 m*M* Neutralized dNTP mix solution.
   c. 10 µL of 2 µ*M* "Upstream" PCR primer, i.e., AP1 or AP3.
   d. 10 µL of 2 µ*M* "Downstream" PCR primer, i.e., AP2 or AP4.
   e. 2 U of *Taq* polymerase.
   f. 1 µg of Template DNA.

   Adjust the reaction vol to 100 µL with nuclease-free H_2O.
2. Overlay the reaction mixture with 50 µL of light liquid paraffin.
3. Subject each DNA sample to 30 cycles of PCR, using a thermal cycler that is programmed such that the initial cycle consists of a 4-min denaturation at 94°C, 2-min annealing period at 55°C, and a 3-min extension period at 72°C. Follow this with 30 cycles of PCR under the following conditions: 94°C (1 min), 55°C (2 min), 72°C (3 min); the last extension reaction at 72°C is prolonged to 10 min.
4. After completion of PCR, remove the mineral oil. Load 5 µL of the reaction together with 250 ng of ϕχ174 Hae III marker in a 1.2% agarose gel and examine the gel after electrophoresis and staining with ethidium bromide to see whether amplification was satisfactory (*see* Note 2).
5. Take 5 µL of each PCR product and make up to 177 µL with H_2O. To denature, add 10 µL of 500 m*M* EDTA and 13 µL of 6*M* NaOH to give a final concentration of 2.5 m*M* EDTA, 0.26*M* NaOH. Stand on ice for 10 min.
6. Soak precut nitrocellulose membrane in H_2O for 10 min. Place the membrane on the dot-blot apparatus and turn on the vacuum for 1 min.
7. Apply 200 µL of 2*M* sodium acetate to the membrane and turn on the vacuum for 1 min, or until all the sample has been aspirated.
8. Apply the denatured samples onto the nitrocellulose membrane and turn on the vacuum again for 1 min.
9. Repeat Step 7.
10. Rinse nitrocellulose membrane in 2× SSC. Blot-dry and bake for 1–2 h at 80°C.

3.3. Preparation of Oligonucleotide Probes (see Note 3)

3.3.1. Radiolabeling of Oligonucleotides

1. Add in the following order:
   a. 15 pmol of oligonucleotide (~100 ng for 19 mer).
   b. H_2O to bring total reaction vol to 10 µL.
   c. 1 µl of 10× kinase buffer.
   d. 1 µL of [γ^{32}P] ATP.
2. Add 2 U of T_4 polynucleotide kinase. Mix and incubate at 37°C for 30 min.
3. If the labeled oligonucleotide is to be separated by gel electrophoresis, add 10 µL of loading buffer and leave on ice.
4. If the labeled oligonucleotide is to be separated by G25-50 spun column chromatography, add 90 µL of TE buffer to bring the total vol to 100 µL and leave on ice.

3.3.2. Separation of Labeled Oligonucleotides by Gel Electrophoresis (see Note 4)

1. Cast a preparative 15% polyacrylamide gel (acrylamide: *bis*-acrylamide = 19:1) of 0.8 mm thickness in 7*M* urea and 1× TBE, using glass plates suitable for the IBI VCV vertical gel electrophoresis box.
2. Preelectrophorese the gel at 25 W for 30 min (w = v × amps).
3. Load the oligonucleotide samples and electrophorese at 25 W until the bromophenol blue (BPB) dye front is at the bottom of the gel.
4. At the end of the run, detach the plates from the tank, lay the gel sandwiched between the two glass plates flat on the bench. Lift a corner of the upper glass plate, leaving the gel attached to the lower plate. Remove the spacers.
5. Cover the gel with cling film and bind the two vertical sides with tape.
6. Place a Kodak X-Omat AR 8" × 10" film over the gel in a dark room and pierce the film over the tape several times with a needle. The needle points act as markers for alignment of the film to the gel.
7. Develop the X-ray film and cut out the labeled bands with a scalpel blade. With the aid of the markers, superimpose the film over the gel; then locate and excise the gel fragments containing the labeled probes. Check that the correct fragments have been excised by reexposing the gel to another X-ray film.
8. Suspend the gel slices in 500 µL of 10 m*M* Tris-HCl, pH 8, 1 m*M* EDTA (TE) overnight at 37°C to elute the labeled oligonucleotides.

9. Remove the eluate and count 5 µL of the eluate in 5 mL of scintillation fluid.

3.3.3. Separation of Labeled Oligonucleotides by Sephadex G25-50 Spun Column Chromatography

1. Remove the plunger from a 1-mL syringe and plug with glass wool. Fill syringe with preswollen Sephadex G25-50 previously equilibrated with TE.
2. Place the 1-mL syringe in a 15-mL Falcon tube so that the fingergrips of the syringe hang from the rim of the tube.
3. Centrifuge at 1000 rpm in the Sorvall RT6000 for 4 min. The Sephadex will pack down. Discard eluate. Add more Sephadex and recentrifuge until the packed vol of the column is 1 ml.
4. Position a "headless" Eppendorf in the Falcon tube. Load 100 µL of TE onto column and centrifuge under identical conditions as for packing column, i.e., 1000 rpm for 4 min. Repeat as necessary until column is equilibrated, i.e., 100 µL is recovered in Eppendorf.
5. Load the labeled ASO that has been made up to 100 µL in TE onto the column and recentrifuge under identical conditions, collecting the sample into a fresh decapped Eppendorf. One hundred microliters of labeled ASO should be recovered.

   It is important not to vary the recentrifugation conditions, since any change will lead to incomplete recovery of the sample.
6. Add 100 µL of TE to the recovered probe to bring the total vol to 200 µL. Take 2 µL, add to 5 mL scintillation fluid, and count (*see* Section 3.4, Step 4 for cpm of ASO required).

3.4. Hybridization and Washing

Generally, all hybridization and washing manipulations of oligonucleotide probes are performed in 1*M* Na^+ conditions, so that the stringency can be altered by the temperature of hybridization and the temperature and time of posthybridization wash. These conditions vary with the length and sequence complexity of each ASO and should have been worked out for each set of oligoprobes using positive and negative DNA controls. It should also be pointed out that some "background" hybridization will be present so that the most important parameter in ASO probe hybridization is the "signal-to-noise" ratio. After the initial wash and exposure, additional washes are usually required to improve the selectivity of hybridization. Thus, where possible, hybridization controls should be included.

1. Remove the dried gel from the Whatman paper backing by floating it on a shallow pan of water. The gel, which is now like a piece of cellophane, will float off the paper after a couple of minutes of gentle shaking.
2. Denature the gel by soaking in 0.5*M* NaOH, 1.5*M* NaCl for 10 min. Rinse with water, and neutralize by soaking in 0.5*M* Tris-HCl, pH 7.5, 1.5*M* NaCl for 10 min.
3. Slip the gel into a polythene bag sealed on three sides (no prehybridization is required). If hybridizing to dot blots, prewet the nitrocellulose membrane in 1× SSC, then slip into a polythene bag and seal on three sides (no prehybridization is required).
4. Hybridize the gel or dot blots with 2×10^6 cpm of 5' end-labeled ASO/mL of 5× SSPE, 0.1% SDS, 100 μg/mL tRNA at the appropriate temperature (usually 5°C below T_d) for a minimum of 2 h.
5. After hybridization, remove the gel or blot and wash in 6× SSC with gentle shaking for 20 min twice at room temperature.
6. Wash for 2–5 min in 6× SSC at the hybridization temperature.
7. Repeat the wash in 6× SSC at room temperature for 1 h.
8. Dry the gel or blot between two sheets of 3MM Whatman paper, wrap in cling film, and autoradiograph with Kodak XAR-5 film between two intensifying screens at –70°C overnight.
9. Repeat Steps 5 and 6, increasing the temperature of wash by 1–2°C as necessary to obtain selectivity of hybridization, and repeat autoradiography, the time of exposure depending on the intensity of the initial signal.
10. Alternative washing procedure. When working with a battery of probes, washing with T-MAC offers a distinct advantage, since all the stringent washes can be performed at the same temperature. After Step 5, wash the blots or gels for 20 min twice at 54°C in 30 mL of T-MAC wash in a polythene bag. Blot the gels dry, wrap in cling film, and expose as in Step 8.

4. Application

4.1. Strategy for Characterization of the Mutant β-Globin Genes

To date, more than 90 mutations are known to cause β-thalassemia *(3)*. Despite this remarkable heterogeneity of molecular lesions, certain observations together with the recently developed recombinant DNA techniques, have made it possible to plan a diagnostic strategy for the identification of the particular molecular defects responsible in individuals with thalassemia. The

important observations were that, among the populations in which β-thalassemia is prevalent, each ethnic group has its own specific types of mutant alleles and that, among each cluster of mutations, there tend to be one or two particularly frequent ones together with a variable number of rare mutations. Therefore, the strategy would be to ascertain the ethnic origin of the individual, and then screen the DNA with a panel of ASO probes for β-thalassemia mutations known to be present in that ethnic group using PCR to amplify specific β-globin gene sequences and dot-blot hybridization. In about 80–90% of the individuals, the mutation should be characterized; the uncharacterized mutations could be determined by direct genomic sequencing of amplified DNA.

4.2. Detection of the IVS-1 Position 1 G-T and the IVS-1 Position 5 G-C β-Thalassemia Mutations

These are two common β-thalassemia mutations among the Asian Indians that are the result of single-base substitutions in the exon-1/intron-1 junction of the β-globin gene. The point mutations are contained in the 5' *Bam*HI 1.9 kb fragment; they occur within four bases of each other and make it possible to use a common normal ASO (β^N), which is completely homologous to the normal coding β-globin gene sequence in this region. In addition, two ASOs specific for the IVS-1 nt I G-T and IVS-1 nt 5 G-C, respectively, were also designed; the sequences of these ASOs are shown in Fig. 1. These ASOs were 5' end-labeled with ^{32}P, purified by polyacrylamide gel electrophoresis and hybridized to total genomic DNA in dried gels or dot blots of amplified DNA, as described in the Methods section. The temperature of hybridization was 53°C, and washing was 53°C in 6× SSC. Figure 2a illustrates the hybridization to DNA in dried gels and Fig. 2b, a dot-blot hybridization.

5. Notes

1. ASO hybridization to total genomic DNA in dried gels suffers from the disadvantage that a minimum of 10 μg of DNA using labeled ASOs separated by polyacrylamide gel electrophoresis is needed to produce a satisfactory signal. This limitation is overcome by amplification of the target sequence.
2. Since the efficiency of amplification in DNA varies from sample to sample, interpretation of dot-blot hybridization of amplified DNAs can be problematical unless comparable amounts of different PCR products are applied in the dot blots. This is best done by comparing small aliquots of the PCR products in an ethidium-bromide stained minigel.

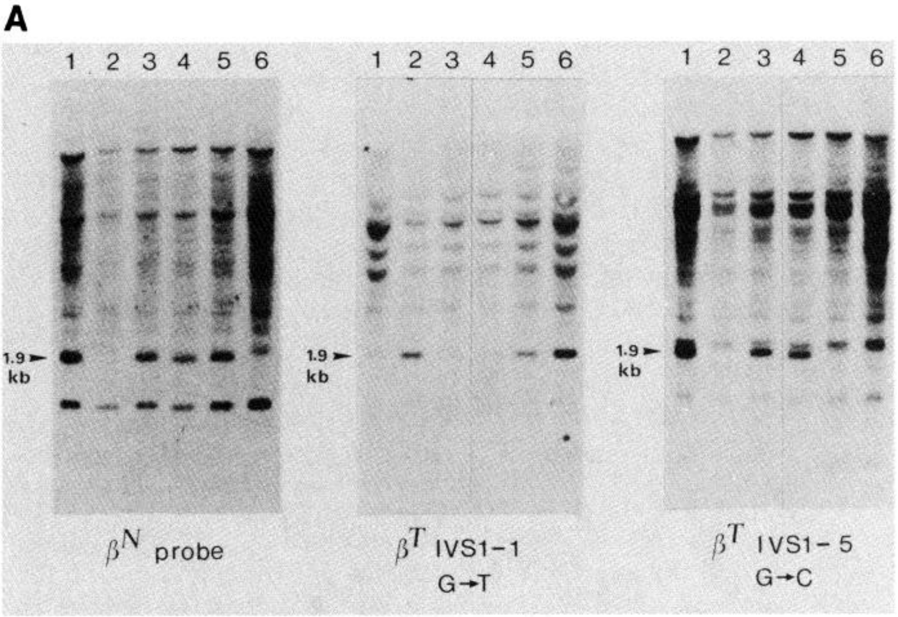


Fig. 2. Detection of the β°IVS1-1 G-T and the β+IVS1-5 G-C mutations using ASOs. Sequences of the oligonucleotide probes are 5'–3'

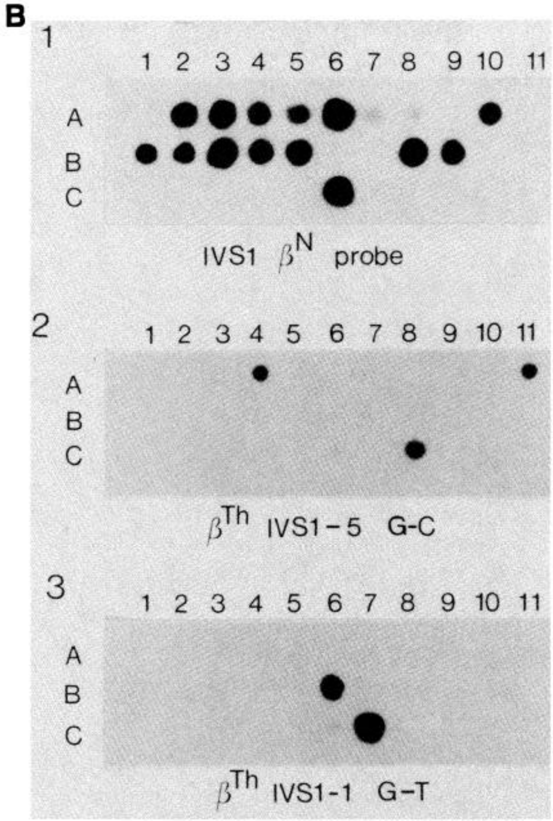
β^N-CTTGATACCAACCTGCCCA

β^{Th} IVS1-1 G-T-CTTGATACCAAACTGCCCA

β^{Th} IVS1-5 G-C-CTTGATAGCAACCTGCCCA

(A) Hybridization to genomic DNA in dried gels. Lanes 1, 3, and 4, positive for both β^N and β^T IVS1-5 probes; lanes 2 and 6, positive for only β^T IVS1-1 probe; 5, positive for both β^N and β^T IVS1-1 probes. Therefore, individuals 1, 3, and 4 are heterozygous for β IVS1-5 mutation, whereas 2 and 6 are homozygous, and 5 is heterozygous for IVS1-1 mutation. **(B)** *(see following page)* Dot-blot hybridization. 1, 2, and 3 are duplicate dot blots of genomic DNA amplified using primers AP1/AP2. A1, A9, B7, B10, B11, and C1–5, C9–11 are blanks. The controls for β^{Th} IVS1-5 G-C and β^{Th} IVS1-1 G-T probes are C8, and B6 and C7, respectively. The results show that A4 is heterozygous and A11 homozygous for β^{Th} IVS1-5 G-C mutation.

3. Enrichment of the target DNA sequence by PCR over the other parts of the genome has increased the sensitivity of detection in dot-blot hybridization, allowing the use of ^{35}S-labeled and nonradioactively labeled probes. ^{35}S-labeled probes can be used for up to 3 mo and nonradioactively labeled probes have a shelf life of 2 yr.
4. ^{32}P-labeled ASOs are best purified by polyacrylamide gel electrophoresis. This method of separation depends on the different mobilities be-



tween the 5' phosphorylated oligonucleotide and the 5' OH oligonucleotide, and therefore, separates the labeled ("hot") oligonucleotide from the unincorporated [γ-^{32}P] ATP as well as from any unlabeled ("cold") oligonucleotides.

5. When working with a battery of ASOs, washing with T-MAC offers a distinct advantage, since this allows the stringency of washing to be controlled as a function of probe length only. Thus, a wash temperature of 52–54°C would be satisfactory for ASOs of 19–22-bp long.

References

1. Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H., and Itakura, K. (1981) The use of synthetic oligonucleotides as hybridisation probes. *Nucleic Acids Res.* **9,** 879–894.

2. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230,** 1350–1354.
3. Thein, S. L. and Weatherall, D. J. (1988) The Thalassaemias, in *Recent Advances in Haematology* (Hoffbrand, A. V., ed.), Churchill Livingstone, UK, pp. 43–74.

CHAPTER 25

Detection of Gene Deletions Using Multiplex Polymerase Chain Reactions

Jeffrey S. Chamberlain, Richard A. Gibbs, Joel E. Ranier, and C. Thomas Caskey

1. Introduction

The polymerase chain reaction (PCR) is a rapid method for the amplification and analysis of DNA sequences, and has greatly simplified the identification of mutations leading to genetic diseases *(1–3)*. The exquisite sensitivity of this method can also be exploited to demonstrate the presence or absence of specific DNA sequences in a sample. This aspect of the procedure has led to the development of assays that can eliminate the need for Southern analysis when screening for DNA deletions that lead to genetic disease. Deletions account for a high frequency of the mutations that have been observed to cause a number of genetic diseases, such as Duchenne/Becker muscular dystrophy (DMD) *(4)*, Lesch-Nyhan syndrome *(5)*, and X-linked ichthyosis *(6)*.

PCR can be used to detect these deletions, and therefore, diagnose the resulting diseases by demonstrating that certain regions of a gene are unable to be amplified. However, the PCR procedure generally is not capable of amplifying regions of DNA larger than a few kilobases (kb) in size, whereas deletions can be highly variable in both size and location within a gene of interest. The ability to multiplex PCR reactions, i.e., amplify a number of sequences simultaneously in a single reaction *(7)*, has led to the development of highly reliable assays that enable large regions of DNA to be efficiently scanned for deletions *(6–9)*. Although multiplex PCR was initially applied to detect hemizygous DNA deletions, the same general procedures can be uti-

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

lized for a variety of purposes, including genetic disease carrier detection, linkage analysis, forensics, multilocus point mutation detection, and DNA library (including yeast artificial chromosome [YAC]) screening.

This chapter provides a detailed description of the use of multiplex PCR for diagnosing deletions that lead to DMD. Such genetic lesions account for between 55 and 65% of all cases of this disease *(4)*. A recent multicenter collaboration among 14 laboratories has found the assay to be reliable for rapid prenatal and postnatal detection of DMD, and to have a detection rate of 82% of all dystrophin gene deletions *(10)*. We will describe how the assay can be performed reliably, discuss potential problems that might be encountered, and how they can be solved or avoided, present a brief overview of additional uses of the technique, and provide generalized concepts for the development of other multiplex PCR assays.

2. Materials

1. Human genomic DNA (250 ng).
2. Disposable gloves.
3. Two sets of microliter pipets (Gilson).
4. dNTPs. We have obtained optimal results using premade 100 m*M* solutions of all four dNTPs purchased from Pharmacia. Equal vol of each dNTP are mixed together and stored as a 25-m*M* stock at –70°C. A working aliquot can be stored for several weeks at –20°C.
5. Dimethylsulfoxide (DMSO) (Aldrich).
6. *Thermus aquaticus (Taq)* DNA polymerase (Amplitaq,® Perkin Elmer Cetus).
7. Oligonucleotide primers (*see* Table 1). These are prepared on an Applied Biosystems Model 380B DNA synthesizer. Primers are deprotected, dried, and stored at –20°C until use; no purification is necessary. For use, the primers are dissolved in 100 µL of autoclaved H_2O, or a vol sufficient to yield an approx 5 mg/mL solution, which is stored at either –20°C or –70°C. Working stocks of each primer are prepared by dilution in H_2O to a concentration of 100 µ*M*, and stored at 4°C. Separate stocks can also be prepared by mixing together equal amounts of each primer and diluting to a 10× (5 µ*M* each primer) concentration, and stored at 4°C. Extreme care should be exercised in pooling primers, as slight errors in the concentration of individual primers can dramatically affect the reliability of the final reaction mixes. Large amounts of individual primers should not be pooled. Fluorescently-labeled primers are prepared as described elsewhere *(11)*, and are stored in H_2O, protected from light at –70°C.

Table 1
Oligonucleotide Primers for Dystrophin Gene Multiplex PCR

	Exon	PCR Primer Sequences, 5'–3'	Amplified Region
A.	Exon 8	F- GTCCTTTACACACTTTACCTGTTGAG	360 bp
		R- GGCCTCATTCTCATGTTCTAATTAG	
B.	Exon 17	F- GACTTTCGATGTTGAGATTACTTTCCC	416 bp
		R- AAGCTTGAGATGCTCTCACCTTTTCC	
C.	Exon 19	F- TTCTACCACATCCCATTTTCTTCCA	459 bp
		R- GATGGCAAAAGTGTTGAGAAAAAGTC	
D.	Exon 44	F- CTTGATCCATATGCTTTTACCTGCA	268 bp
		R- TCCATCACCCTTCAGAACCTGATCT	
E.	Exon 45	F- AAACATGGAACATCCTTGTGGGGAC	547 bp
		R- CATTCCTATTAGATCTGTCGCCCTAC	
F.	Exon 48	F- TTGAATACATTGGTTAAATCCCAACATG	506 bp
		R- CCTGAATAAAGTCTTCCTTACCACAC	
G.	Exon 12	F- GATAGTGGGCTTTACTTACATCCTTC	331 bp
		R- GAAAGCACGCAACATAAGATACACCT	
H.	Exon 51	F- GAAATTGGCTCTTTAGCTTGTGTTTC	388 bp
		R- GGAGAGTAAAGTGATTGGTGGAAAATC	
I.	Exon 4	F- TTGTCGGTCTCCTGCTGGTCAGTG	196 bp
		R- CAAAGCCCTCACTCAAACATGAAGC	

8. NuSieve GTG agarose (FMC Bioproducts).
9. Ethidium bromide.
10. Paraffin oil (light mineral oil).
11. 5× *Taq* polymerase buffer: 83 m*M* $(NH_4)_2SO_4$; 335 m*M* Tris-HCl, pH 8.3; 33.5 m*M* $MgCl_2$; 50 m*M* β-mercaptoethanol; 850 μ*M* bovine serum albumin (BSA); and 34 μ*M* EDTA. The polymerase buffer is generally stored in 1-mL aliquots at –70°C. A working stock of one tube may be kept at –20°C. The buffer is mixed together from premade stocks (autoclaved) of each salt at 1*M* (or 0.5*M*) concentration. Mercaptoethanol is added from a 1*M* solution stored at 4°C, and BSA (nuclease-free) is kept at –20°C as a 50 mg/mL solution. The buffer is stable at least 3 mo at –70°C.

12. 0.5-mL Microfuge tubes.
13. Thermocycler (Perkin Elmer Cetus).
14. DNA gel electrophoresis apparatus (Model MPH, International Biotechnologies, Inc.).
15. 10× Electrophoresis buffer (10× TBE): 900 m*M* Tris-base, 900 m*M* boric acid, 1 m*M* EDTA.
16. Template DNA. Human genomic DNA can be prepared by a variety of methods as long as care is taken to avoid contamination of the DNA by plasmids, PCR reaction products, or other human DNA samples (*see* Section 4.1).

    We generally prepare all samples on an Applied Biosystems Model 340A DNA extractor. DNA prepared on this machine has always proved to be of sufficient quality for PCR and has not resulted in any cross-contamination of samples. The source of DNA can vary considerably, depending on the type of analysis being performed. Blood drawn in the presence of either EDTA or heparin has always yielded good results, although heparin-treated blood can occasionally yield DNA refractory to amplification when prepared by manual extraction procedures. Care should be taken to ensure that neither heparin nor EDTA remain in the final DNA preparation at levels sufficient to either inhibit *Taq* polymerase or interfere with effective Mg^{2+} concentrations. Other sources of DNA include amniotic fluid cells, chorionic villus specimens (CVS), cultured lymphoblasts, and biopsy materials (including paraffin-embedded samples). CVS tissue should be microscopically dissected of maternal decidual tissue to prevent false-positive amplification of maternal DNA during a prenatal diagnosis *(7)*.
17. Reaction mixes. Premade aliquots of the reaction mixes are prepared in bulk and stored at –70°C in either 45-µL individual "kits" (in 0.5 mL microfuge tubes), or as 1-mL stocks (in screw capped microfuge tubes). If the larger aliquots are used, they should be stored at –20°C after initial thawing, and used as quickly as possible. Repeated freeze-thawing should be kept to an absolute minimum. Preparation of aliquoted kits from a large batch of reagents ensures greater sample to sample consistency, and allows each batch to be quality controlled to guarantee effectiveness and lack of contamination by exogenous DNA. Kits are prepared as follows:

Add the following to a 13 × 100-mm sterile polypropylene tube, mix gently after each addition: 2.7 mL H_2O, 1 mL 5× *Taq* polymerase buffer, 300 µL dNTP stock (25 m*M* each), 2.5 nmol of each oligonucleotide primer (500

μL 10× stock), and 500 μL of DMSO. The reaction mix can then be aliquoted and stored. Reaction mixes are stable at least 6 mo at –70°C. Several aliquots should be tested immediately with positive and negative controls to ensure quality *(see below)*.

3. Methods

3.1. Running the Reactions

1. Thaw an individual 45-μL reaction kit, or aliquot 45 μL from a larger pool into 0.5-mL microfuge tubes.
2. Add 250 ng of template DNA. (Dilute DNA in H_2O to a final concentration of between 50 and 250 ng/μL, so that the DNA may be added to the microfuge tube in a vol of 5 μL or less.) Add H_2O to a final vol of 50 μL.
3. Add 5 U of *Taq* polymerase and mix gently.
4. Add 30 μL of paraffin oil, centrifuge for 5 s.
5. Place samples in an automatic thermocycler. For the Perkin Elmer Cetus machine, cycle as follows:.
   a. 94°C × 6 min.
   b. 94°C × 30 s.
   c. 53°C × 30 s.
   d. 65°C × 4 min.
   e. Repeat steps b–d for a total of 23–25 cycles.
   f. 65°C × 7 min.
   g. Store at 4°C until analysis (up to 2 mo).

3.2. Analyzing the Reactions

1. Prepare a 90-mL agarose gel. For optimal resolution of the tightly spaced amplification products, a 3% NuSieve agarose gel is recommended. Alternatively, a 1.5% conventional agarose gel can also be utilized, but this will not produce a very clear final result. NuSieve agarose gels can be tricky to handle, and the brittleness of the gels can be reduced by the addition of 10% conventional agarose (e.g., LE agarose, FMC Bioproducts). NuSieve agarose should be stirred after adding the running buffer (1X TBE) to eliminate trapped air. To dissolve in a microwave oven, heat for 2 min on MEDIUM (70% full power), swirl gently, and then bring solution to a boil for 30 s on HIGH or until all agarose has dissolved. Cool the solution to 55°C, add ethidium bromide to 0.5 μg/mL, and pour onto a gel tray (precooled) at 4°C.
2. Remove the sample from the cooled reaction with a Gilson P20 microliter pipet or equivalent, pipeting from below the layer of oil, and wipe

any adhering oil from the pipet tip with a kimwipe, as the oil will interfere with the migration of the DNA through the gel. Alternatively, the oil can be removed by adding 50 µL of chloroform, mixing and centrifuging for 10 s. The oil dissolves in the $CHCl_3$ (bottom layer).

3. Load 15 µL of the reaction product on the gel and electrophrese at 100 V (3.7 V/cm) for 2 h. We have obtained optimal resolution by utilizing an IBI MPF gel electrophoresis system, and electrophoresing 15 µL of the reaction products at 100 V (3.7 V/cm) for 2 h.
4. Record the final results by photographing the gel with a polaroid camera and a UV transilluminator.

To ensure that the reaction results are reliable, both positive and negative controls must be performed in parallel with each analysis. A positive control consists of human DNA known to carry a normal dystrophin gene, and a negative control consists of a reaction to which no human DNA is added. Additional controls can be performed by amplifying DNA samples that contain partial dystrophin gene deletions that have been previously delineated via Southern analysis.

An example of the use of multiplex PCR for detecting deletions in the dystrophin gene is shown in Fig. 1. At the top of the figure is a schematic illustration of the gene, indicating the relative location of the nine exon-containing regions that are coamplified in the reactions. Additional details on each of these nine regions are presented in Table 1. Also displayed in Fig. 1 is a photograph of a gel through which several completed reactions were electrophoresed. Lane A contains a sample that did not display a deletion, and all nine amplification products are clearly evident. Lanes B–E contain samples that displayed various partial deletions of the dystrophin gene, whereas the sample in lane F displayed a complete deletion of all nine of the regions analyzed. Samples that display a complete deletion are derived from a very low percentage of all DMD patients and should be reanalyzed via Southern analysis to confirm the results. Alternatively, an additional primer set could be added to the reactions as an internal positive control to ensure that amplification was not inhibited.

3.3. Automating the Assay

Multiplex PCR can be applied to a wide variety of both research and clinical applications *(see below)*, and its use is facilitated by the simplicity and rapidity of the assays. The utility of the method can be further augmented by automation of various steps of the procedure. Currently, the reactions are set up in a kit form, enabling greater sample to sample consistency, as well as simplifying the assay by eliminating the need for preparation of reagents for each individual analysis. Amplification is performed on automated thermo-

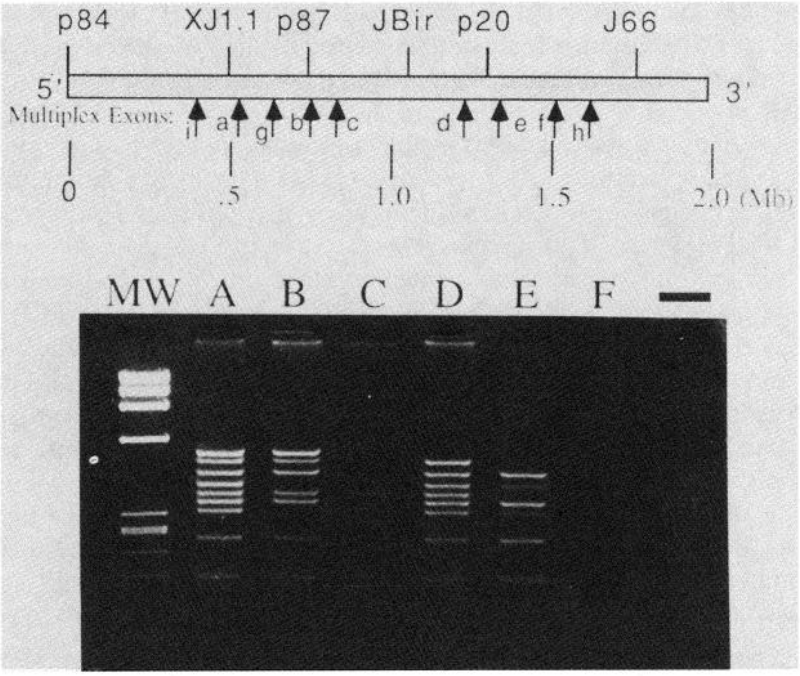


Fig. 1. Multiplex PCR at the dystrophin gene. Top: Schematic illustration of the DMD gene indicating the relative location of the nine exon-containing DNA segments amplified with this procedure (arrows, a–i). Also shown are the approximate locations of several RFLP-detecting genomic probes frequently used for haplotyping. The exon contained within each amplified region is indicated in Table 1. Bottom: Detection of deletions in the DNA of DMD males. M.W: *Hae*III digested φX174 DNA mol wt standard. A–F display the results of multiplex amplification from the DNA of six unrelated male DMD patients. The sample in (A) does not display a deletion (normal pattern of amplification), the samples in (B–E) display a deletion of one or more of the nine regions, and the sample in F was deleted far all nine regions. (–) was a negative control in which no template DNA was added to the reaction. PCRs were performed and analyzed as described in the text. Shown is a photograph of a 3% NuSieve agarose gel through which 15 μL of each reaction was electrophoresed. From top to bottom, the amplified fragments correspond to exons e, f, c, b, h, a, g, d, and i, respectively. Reprinted from *(8)* with permission.

cyclers, and the products are analyzed via manual gel electrophoresis in the presence of ethidium bromide. We have recently developed a modification of the assay that permits the reaction products to be automatically analyzed on an Applied Biosystems Model 370A DNA sequencer (*see also* Chapter 12). For this procedure, one member of each of the pairs of PCR primers is fluorescently end-labeled with the fluorescein dye FAM *(11)* and either uti-

lized in place of the original PCR primer or added to a reaction mix as a percentage of the total primer concentration (i.e., 10% fluorescent, 90% unlabeled primer). Multiplex PCR is then performed as described above, except that the reaction products are analyzed on the Model 370A. This modification of the assay presents several advantages over the use of manual gel electrophoresis. First, the fluorescent dyes eliminate the need for ethidium bromide, a powerful mutagen, while actually increasing the sensitivity of detection. This latter feature enables fewer cycles of PCR to be performed, which reduces the risk of false-positive amplification from contaminating maternal or exogenous DNA *(7)*, reduces the time necessary for PCR, and increases the linearity of fragment coamplification, which facilitates using the method for quantitation of gene dosage *(12; see below)*. Second, analysis of reactions on the 370A eliminates the need to monitor the gels or photograph the results. Instead, the results are stored automatically in a computer and can be recalled at a later time. Figure 2 displays an example of two samples that were amplified from six of the nine primer pairs displayed in Table 1. In both cases, one member of each primer pair was replaced with the corresponding FAM-labeled primer, DNA was amplified for 19 cycles of PCR, and analyzed on the Model 370A. One of the samples displayed the normal amplification pattern, whereas the second displayed a deletion of three of the six regions analyzed.

4. Notes

Through extensive development and testing of this method, we have observed that virtually all problems encountered can be traced to one of the following causes.

1. Contamination by maternal or exogenous DNA. Failure to achieve amplification of a DNA fragment is the only indication that a mutation has been identified, therefore, false-positive amplification from maternal or contaminating exogenous DNA could theoretically mask a deletion and lead to a misdiagnosis. We have observed through reconstruction experiments that levels of maternal DNA (e.g., from amniotic fluid cells or decidual tissue in a CVS) at up to 5% of the total will not lead to false positive amplification as long as the reactions do not approach saturation *(7)*. Typically, the number of PCR cycles performed is kept to the minimum necessary to produce a clear signal in the final gel analysis (generally 23–25 for manual gels, 19–23 for fluorescent assays). The relative amplification of each of the nine DNA fragments remains essentially constant for approx 23 cycles of PCR, but as more cycles are performed, the ratio can change dramatically as some fragments stop amplifying

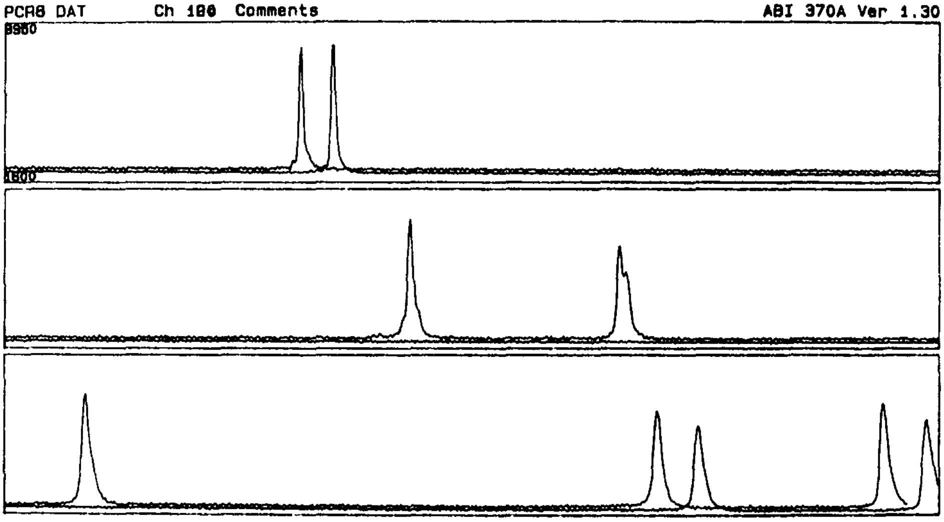


Fig. 2. Automated analysis of fluorescent multiplex PCR products. Six PCR primer sets [A,B,E,F,G,I, (Table 1 and Fig. 1); one member of each set labeled with FAM (*11,12*)], were used for multiplex PCR analysis of two DMD patient DNAs. After 19 cycles of PCR, 3 µL of each reaction was electrophoresed on an ABI 370A. Shown is the computer generated display of the relative fluorescence observed in each sample plotted against time of electrophoresis (upper left, smallest DNA fragments, bottom right, largest). Both sample results were printed together, slightly offset. All six PCR products were observed with the first sample (left-hand member of each pair of peaks); three products were missing from the second sample, revealing the presence of a deletion in this patient's DNA. The original display is in color, allowing clear interpretation of which peaks correspond to which sample.

and others (including possible contaminating bands) continue to accumulate exponentially. These observations have also led to the development of quantitative multiplex PCR assays *(12,13)*. Uniform coamplification of all nine of the regions indicated in Table 1 can be achieved with FAM-labeled primers, and as long as the reactions are performed for between 16 and 23 cycles of PCR, the relative amplification of each region can be compared and used to detect heterozygous or homozygous dystrophin gene deletions and duplications in both carrier females and affected males *(10–13)*.

In addition to the number of cycles of PCR, care should also be taken in preparation of the template DNAs. Maternal decidual tissue should be microscopically dissected from CVS tissue prior to extraction of DNA. Equipment that has been in contact with either prior reaction products

or cloned DNA complementary to any of the regions being amplified should not be used to prepare template DNA. Disposable gloves should always be worn when performing PCR assays. Separate pipetors must be used to sample the amplified reactions from those used to prepare template DNAs or to mix together the reaction components. In addition, the preparation and analysis stages of the reaction should be physically separated. Amplified reactions are opened and aliquots removed for analysis at a separate location than where the reactions are initiated. These latter precautions are critical to prevent trace quantities of prior reaction products from serving as efficient template for future reactions. To eliminate or preclude the possibility of contaminated pipetors, they may be effectively cleaned by soaking the barrel of the pipetman in 0.25*N* HCl for 30 min, then again with 0.5*N* NaOH for 30 min, and then rinsing well with H_2O. By following these precautions, more than 700 DNA samples have been analyzed via multiplex PCR without encountering a problem related to false-positive amplification *(10)*.

2. The reactions have been optimized to produce clearly visible results from 23 cycles of PCR starting with 250 ng of template DNA. The amount of template DNA added to the reactions will therefore affect the final results. Too little DNA may not produce a detectable signal after 23 cycles of PCR. In such cases, the reaction can be returned to the thermocycler for a few additional cycles (no additional enzyme is required and samples can be reamplified up to 1 wk after the initial amplification). However, performing additional cycles increases the possibility of false-positives, as indicated above. Too much template DNA added to the reactions can distort the ratio of the amplified fragments, complicating interpretation of the results.
3. The parameters of the PCR reaction and the type of thermocycler used can drastically alter the results of an assay. The conditions listed above for the PCR cycle profiles have been optimized for the Perkin Elmer Cetus machine using 'step file' functions. We have observed that different machines made by the same manufacturer may need slightly altered annealing temperatures for optimal results. We recommend that the conditions listed above be used initially, and if unsatisfactory results are obtained, the annealing temperature can be raised 1–2°C (if extra bands are produced) or lowered a few degrees (if not all nine bands are produced—this is usually apparent by failure to obtain the fragment corresponding to exon 48 [Table 1], which has the primer set with the lowest melting temperature). Other manufacturer's machines may require additional modifications in the settings. Almost all problems resulting

from machine variations can be corrected by adjusting the annealing temperature. Annealing should be performed at the highest possible temperature for the shortest amount of time (usually 30 s).

4. For as yet undetermined reasons, we have observed that approx 1% of the reactions will produce a false negative amplification pattern. In each case, the pattern obtained indicated that the two largest fragments had not amplified (suggesting a deletion of dystrophin exons 45 and 48—Table 1). Also, in each of these cases, when the assay was repeated the correct pattern was obtained. For these reasons, any PCR reaction displaying a deletion should either be repeated, or preferably, the deletion should be confirmed via Southern analysis with a dystrophin cDNA subclone corresponding to a region included within or overlapping the deletion. In addition, although we have not yet encountered such a situation, one should ensure that an observed deletion corresponds to consecutive regions of the dystrophin gene. To date there have been no reports of more than a single deletion within the dystrophin gene of any one individual.

4.1. Additional Applications

This chapter has focused on the use of multiplex PCR for detecting deletions in the human dystrophin gene. However, the basic concept of multiplexing PCRs has already been applied to several other disease genes. Multiple primer sets have been used to coamplify each of the exons of the hypoxanthine phosphoribosyltransferase (HPRT) gene. This multiplex PCR has been used to detect both deletions and point-mutations leading to Lesch-Nyhan syndrome *(9)*. A third multiplex PCR assay has been reported that detects 100% of all deletions leading to steroid sulfatase deficiency (X-linked ichthyosis) *(6)*. This latter assay requires only two separate primer pairs, but has the added advantage of including one of the dystrophin gene PCR primer pairs to serve as an internal control for amplification. In addition, we have recently developed a multiplex PCR assay that enables coamplification of a number of genes involved in the most common genetic diseases that arise from point mutations *(14)*.

As several of the applications listed above indicate, many of the newer multiplex PCR assays are not designed to detect deletions, but instead, are utilized to amplify multiple regions of the genome for either point mutation or polymorphism detection. One such assay simultaneously amplifies multiple polymorphic bases and short tandem repeats within the human dystrophin gene *(15)*. This assay should complement the existing dystrophin PCR test, enabling linkage analysis to be carried out via PCR for those DMD families that do not display a genomic DNA rearrangement. The ability to

multiplex PCRs should continue to find wide application to both genetic diseases as well as any assay requiring that multiple regions of DNA be analyzed simultaneously. We have recently begun using such assays as a rapid method to screen genomic DNA libraries (particularly YAC libraries), to perform forensic identification of DNA samples, to perform linkage analysis for human genome mapping, and to quantitate gene dosages. Each of the newer applications has been aided by experience derived from the early assays, and has led to several principles for the development of any multiplex PCR.

Initially, highly specific DNA sequence information must be obtained. For example, the X-linked STS gene displays very high homology with a Y-linked pseudogene, which complicated efforts to obtain X-specific amplification *(6)*. The locations of PCR priming sites need to be chosen while considering the entire multiplex reaction. Flexibility in choosing the size of the regions to be amplified facilitates obtaining multiple reaction products that can be resolved on agarose gels. Successful primers have generally been 23–28 bp in length, sufficient to permit high stringency annealing, and thus highly specific amplification. The percentage of Gs and Cs within a primer sequence is also important. Primers with 40–60% G/C contents generally work well, and all primers in a multiplex reaction should display similar melting temperatures. Primers with low G/C contents often amplify poorly, whereas too high a G/C content can lead to the appearance of spurious amplification products.

Occasionally, primers that have worked well individually have led to spurious amplification products (or no amplification at all) when multiplexed. Frequently, any such problem in a mixture of primers can be traced to only one or two of the primers, which can then be replaced with alternative primers (often synthesizing a new primer displaced from the original primer by a few bases will solve the problem). However, before going to the time and expense of synthesizing new primers, several variables in the assay should be altered in an attempt to generate working reactions. Important parameters that can enable PCRs to be multiplexed are:

1. The amount of enzyme—more fragments require more enzyme.
2. The ratio of the primers—although the assay for dystrophin deletions uses equimolar ratios of all 18 primers, the HPRT assay was improved by reducing the concentration of some of the primers relative to others.
3. The annealing temperature must be reoptimized for any new set of primers.
4. The polymerase extension time generally must be 2–4 times longer for a mixture of primers than for any single pair that is in the reaction.
5. Mg^{2+} and dNTP concentrations can affect the reliability of the assays.

Mg^{2+} levels must be balanced against both the individual requirements of each primer as well as the amount of dNTPs present in a reaction. Separate combinations of primers require different conditions for multiplex amplification, and it will probably be necessary to optimize reaction conditions for any given set of oligonucleotides. Finally, as more primer sets are added to a multiplex PCR, the permissive reaction conditions generally become increasingly less flexible.

Acknowledgments

We thank Phi-Nga Nguyen, Nancy Farwell, Donna Muzny, and Andrew Civitello for excellent technical assistance. This work was supported by a Task Force on Genetics grant from the Muscular Dystrophy Association, and by the Texas Advanced Technology Program under Grant 3034. JSC was supported by a postdoctoral fellowship from the Muscular Dystrophy Association. RAG is a recipient of the Muscular Dystrophy Association's Robert G. Sampson Distinguished Research Fellowship. CTC is a Howard Hughes Medical Institute Investigator.

References

1. Erlich, H. A (ed) (1989) *PCR Technology Principles and Applications of DNA Amplification.* Stockton, New York
2. Gibbs, R A. and Chamberlain, J S (1989) The polymerase chain reaction A meeting report *Genes Dev* **3,** 1095–1098
3. Erlich, H. A., Gibbs, R A., and Kazazian, H. H., Jr. (eds.) (1989) *The Polymerase Chain Reaction: Current Communications in Molecular Biology.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
4. Chamberlain, J. S. and Caskey, C. T. (1990) Duchenne muscular dystrophy, in *Current Neurology,* vol. 10, Chapter 2. Yearbook Medical, Chicago, IL
5. Stout, J T. and Caskey, C. T (1985) HPRT: Gene structure, expression and mutation. *Annu. Rev. Genet.* **19,** 127–148.
6. Ballabio, A., Ranier, J E., Chamberlain, J. S , Zollo, M., and Caskey, C. T (1990) Screening for steroid sulfatase (STS) gene deletions via multiplex DNA amplification *Hum Genet.* **84,** 571–573.
7. Chamberlain, J. S , Gibbs, R A , Ranier, J. E., Nguyen, P. N., and Caskey, C. T (1988) Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Res.* **16,** 11141–11156.
8. Chamberlain, J. S., Gibbs, R A , Ranier, J. E., and Caskey, C. T. (1989) Multiplex PCR for the diagnosis of Duchenne muscular dystrophy, in *PCR Protocols: A Guide to Methods and Applications* (Innis, M., Gelfand, D., Sninski, J., and White, T., eds.), Academic, Orlando, FL, pp 272–281
9. Gibbs, R A., Nguyen, P N , Edwards, A. O., Civitello, A , and Caskey, C. T (1990) Multiplex DNA deletion detection and exon sequencing of the hypoxanthine phosphoribosyltransferase gene in Lesch-Nyhan families. *Genomics* **7,** 235–244.

10. Chamberlain, J. S., Ranier, J. E., Caskey, C. T., et al. (1991) Results of a multicenter collaboration of the efficiency and effectiveness of multiplex PCR for diagnosis of Duchenne muscular dystrophy. Submitted to *N. Engl. J. Med.*
11. Gibbs, R. A., Nguyen, P. N., McBride, L. J., Koepf, S. M., and Caskey, C. T. (1989) Identification of mutations leading to the Lesch-Nyhan syndrome by automated direct DNA sequencing of in vitro amplified cDNA. *Proc. Natl. Acad. Sci. USA* **86,** 1919–1923.
12. Chamberlain, J. S., Ranier, J. E., Gibbs, R. A., Farwell, N. F., McBride, L. J., Madden, D., and Caskey, C. T. (1991) The use of PCR for diagnosis of mutations in the mouse and human dystrophin genes Submitted to *J. Cell Biochem*
13. Fenwick, R., Chamberlain, J. S., Ranier, J. E., and Caskey, C. T. Unpublished observations.
14. Grompe, M., Chamberlain, J. S , Gibbs, R. A., and Caskey, C. T Unpublished observations.
15. Chamberlain, J. S , Gibbs, R. A., Ranier, J. E., and Caskey, C. T. (1989) An integrated approach to Duchenne muscular dystrophy diagnosis via multiplex polymerase chain reaction. *Am. J. Hum. Genet.* **45,** A134.

CHAPTER 26

Application of Pulsed-Field Gel Electrophoresis to Genetic Diagnosis

Johan T. den Dunnen and Gert-Jan B. van Ommen

1. Introduction

Hereditary diseases *(1)* are diseases that are passed on from one generation to the next. They are caused by one or more genetic defects in consequence of point mutations, small insertions and deletions, or chromosomal rearrangements, notably deletions, duplications, inversions, insertions, and translocations.

Any method for the detection of point mutations requires a nucleotide-by-nucleotide comparison of a normal and a defective gene. Therefore, the detection of point mutations has until recently been difficult and laborious. However, the advent of polymerase chain reaction (PCR) techniques has dramatically altered the prospect of this field of research (*see* Chapters 1–14, this volume).

In principle, the detection of chromosomal rearrangements should be easier. They produce size differences when chromosomes or DNA fragments of normal and diseased persons are compared. The size differences can be detected either cytogenetically or by electrophoresis and blotting. Light-microscopic cytogenetics presently allows the detection of only those chromosomal rearrangements involving at least 5–10 million bp (Mbp) of DNA. The standard technique used for the separation of DNA fragments, agarose gel

From *Methods in Molecular Biology, Vol. 9. Protocols in Human Molecular Genetics*
Edited by: C Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

electrophoresis, is capable of separating only fragments up to 30 kilobase pairs (30 kbp). This limits the detection of rearrangements to within 30 kbp of a specific site.

Pulsed-field gel electrophoresis (PFGE) is a newly developed technique that enables the separation of DNA fragments up to 6.0 Mbp (*see* Chapter 17). This extends the detection "window" of chromosomal rearrangements by two orders of magnitude and thus enormously increases the chance to detect abnormalities. Furthermore, the diploidy of somatic cells hinders the detection of haploid loss or duplication of sequences by Southern blotting, since this needs to be done by dosage comparisons, producing 1/2 or 3/2 ratios for deletions and duplications, respectively. In a PFGE analysis, deletions and duplications are detected as differences in fragment sizes.

The large potential of PFGE to address diagnostic questions is demonstrated by its application in the study of the Duchenne Muscular Dystrophy (DMD) gene *(2–8)*. DMD is an X-linked progressive muscle-wasting disorder that affects one in 3500 boys and ultimately leads to death of the patients in early adulthood *(9)*. Recently, the "reverse genetics" approach has led to the identification of the underlying genetic defect *(10,11)*. The DMD gene turns out to encode a 14-kbp mRNA, which is translated into a 427-kDa sarcolemma-associated protein *(11)*, called dystrophin, the exact function of which is not yet fully understood (for review, *see* ref. *12*).

The most remarkable feature of the gene is its enormous size; it measures 2.3 Mbp *(7,8)*. This size was established by using PFGE analysis to construct a physical map of the DMD region. The first map was made using genomic probes encompassing the gene *(2,3,13)*; subsequently intragenic probes were mapped *(5,14)*, and finally the gene boundaries were localized using the cDNA *(7,8)*.

The size of the gene, and the detection of deletions in about 10% of the patients by using the intragenic probe pERT87 (DXS164), prompted us to use PFGE to screen DMD patients for chromosomal aberrations. This rapidly resulted in our discovery that, in over 50% of the cases, large deletions or duplications were responsible for the disease *(4,15)*. This finding was confirmed independently by Southern-blot analysis with cDNA probes *(16)*. More extensive analysis has shown that 60–70% of DMD patients carry deletions or duplications *(7,16–18)*.

This chapter describes the application of PFGE to study chromosomal rearrangements in a specific genomic region. First, it discusses which criteria justify an initiation of this type of study. Second, with DMD as an example, it shows its practical application, highlighting both its unique possibilities and its limitations.

2. Strategy

2.1. Types of Chromosomal Aberrations

Figure 1 shows the schematic result of a conventional Southern blot and of a PFGE analysis of each different type of rearrangement when these affect either autosomal sequences or X-specific sequences in female DNA.

PFGE analysis does not require the comparison of hybridization intensities necessary to detect deletions and duplications on a conventional Southern blot (compare Figs. 1B and 1C), since both types of rearrangement produce altered fragments. Deletions involving a rare-cutter site create an abnormal "fusion" fragment, detectable by probes that normally detect different fragments. Translocations are especially hard to detect by any method other than PFGE analysis. Using PFGE, translocations result in two abnormal fragments, detectable with probes from the opposite ends of the original fragments. Each translocation junction fragment is specific for one derivative chromosome (Fig. 1C). Inversions, equally hard to detect in conventional electrophoresis, mostly show two altered fragments in PFGE analysis. For insertions, the result of a PFGE analysis is comparable to that for duplications, i.e., a single fragment of increased size. On conventional blots, insertions will be detected only when they are located within 10–20 kbp of a probe.

The picture shown in Fig. 1 is an oversimplification of the practical situation. The results may be more complex when deletions involve the complete probe, insertions contain a new rare-cutter site, inversions are completely contained within one fragment, translocation or inversion breakpoints are very close to the restriction sites, or the rearrangement itself is a combination of several types. However, most of these problems can be solved by using different restriction enzymes.

Furthermore, the results may be obscured by partial digestion *(see below)* or by naturally occurring restriction fragment length polymorphisms (RFLPs). These not only involve point mutations, leading to loss or gain of a restriction site, but also, notably, deletions, duplications, and insertions not related to the disease phenotype.

Therefore, before a definite conclusion can be drawn about the rearrangement that underlies the PFGE abnormality detected, one should be sure that:

1. The involvement of RFLPs can be excluded by a PFGE analysis of DNA of a set of control individuals,
2. Probes are used from both ends of the altered fragment(s) (probes A and D in Fig. 1), and
3. The result is verified with a second restriction enzyme.

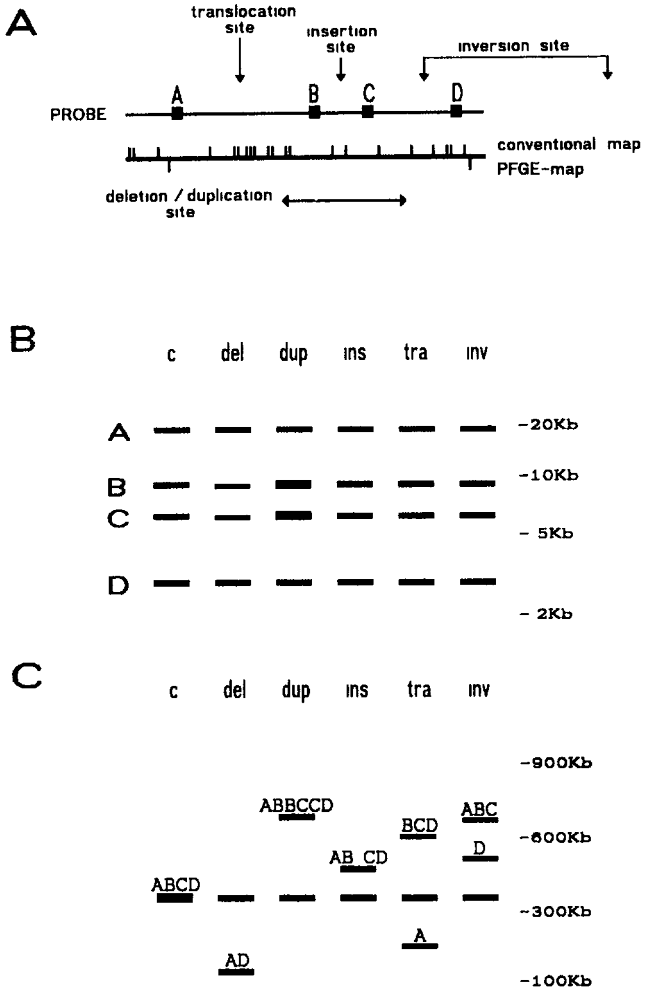


Fig. 1. Chromosomal aberrations and their effect on conventional and PFGE analysis in female DNA: **(A)** Physical map. Black squares show the location of the probes. The site and extent of imaginary mutations are indicated. **(B)** Hybridization pattern of a conventional agarose analysis. **(C)** Hybridization pattern of a PFGE analysis. The samples used each carry one of the rearrangements depicted in A; normal (c), deletion (del), duplication (dup), insertion (ins), translocation (tra), and inversion (inv). Sizes are indicated on the right. Probes detecting the fragments shown are given on the left or above altered PFGE fragments.

If possible, the result should be confirmed by conventional Southern-blot analysis using cDNA probes of the gene itself. Usually, once one knows what to look for, this analysis becomes informative as well.

2.2. Physical Map

The basis for the application of PFGE to analyze a specific genomic region is the availability of a good physical map of that region. This map is constructed using a combination of single and double digestions with several rare-cutter restriction endonucleases. All available probes should be tested, and finally the borders of the target gene should be defined as precisely as possible. Subsequently, one or two rare-cutters are selected that cover the whole region with one, or a few, clearly detectable restriction fragment(s). This selection is combined with the choice of specific electrophoretic conditions that optimize the detection of size differences in the fragment(s) under study before the analysis can be started.

3. Examples

3.1. The Dystrophin Gene: Physical Map

PFGE analysis of the DMD gene revealed two main features: First, the gene is extremely large *(7,8,14)*, so it could not be contained within one restriction fragment; second, most restriction enzymes failed to give clearly detectable fragments for all parts of the gene *(14)*.

SfiI appeared to be the best restriction enzyme for the analysis (*see* refs. *4, 6,* and *8,* and Fig. 2). A pilot study, using a limited set of control persons, did not reveal RFLPs in the SfiI map *(4,8)*. The SfiI map of the DMD gene shows five clearly discernible restriction fragments in the size range of 200–800 kbp. Three of these fragments contain partially digestible sites, which brings the total of intragenic sites to seven (Fig. 2).

3.2. Chromosomal Rearrangements

DMD is an X-linked disease. Consequently, the analysis of male patients with intragenic DNA probes lights up sequences from only one chromosome (compare with Fig. 2).

An example of a DMD deletion is given by patient DL23.4 (Fig. 3A). The SfiI fragments CD and FI have a normal size. However, fragment EF is altered; it is 200 kbp smaller, indicating a deletion in this part of the DMD gene. Analysis with the dystrophin cDNA confirms this; four exons are missing *(7)*.

Figure 3B shows a more complex pattern resulting from the partial digestibility of some restriction sites. Knowledge of the normal physical map, however, allows this pattern to be explained by a 130-kbp deletion removing SfiI site F and producing several abnormal partial fragments. Thus, although the partial digestion complicates the emerging picture, at the same time it pinpoints the rearrangement to a precise location. In fact, at an earlier stage,

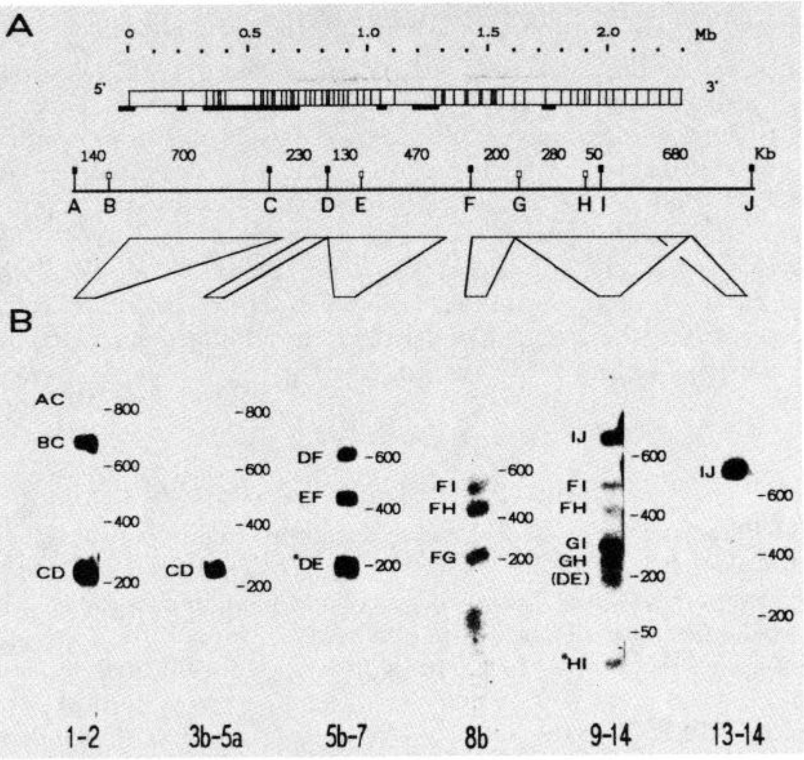


Fig. 2. Megabase map of the DMD gene. **(A)** SfiI physical map of the DMD region (bottom line) showing partially (open boxes) or fully (closed boxes) digestible SfiI sites. Individual sites are marked with letters *(4)* and fragment lengths are indicated in kb. The top line shows the localization of exon-containing genomic HindIII fragments (vertical bar) in relation to the SfiI map *(4,7)*. Heavy lines at the bottom of exon-containing fragments show the extent of cloned regions. **(B)** cDNA hybridizations to FIGE gels in relation to the megabase map. SfiI-digested DNAs were hybridized to specific cDNA subclones *(7,16)*; bottom. Letter symbols (left) mark each fragment detected (*see* A), asterisks mark abnormally migrating fragments, and brackets mark a signal of a previous hybridization not removed fully after stripping the filter. Fragments sizes are indicated at the right (in kb).

partial digestion combined with deletion data greatly assisted in constructing the DMD megabase map *(5)*.

A duplication as the chromosomal rearrangement causing DMD is found in patient DL150.5 (Fig. 3C). SfiI fragments CD and FI have normal sizes but fragment EF is 200 kbp larger than normal. Furthermore, the hybridization signal obtained with probe JBir is stronger than that obtained with other probes. This suggests a duplication of sequences, including JBir. A cDNA hybridization confirmed a duplication involving exonic sequences *(7)*.

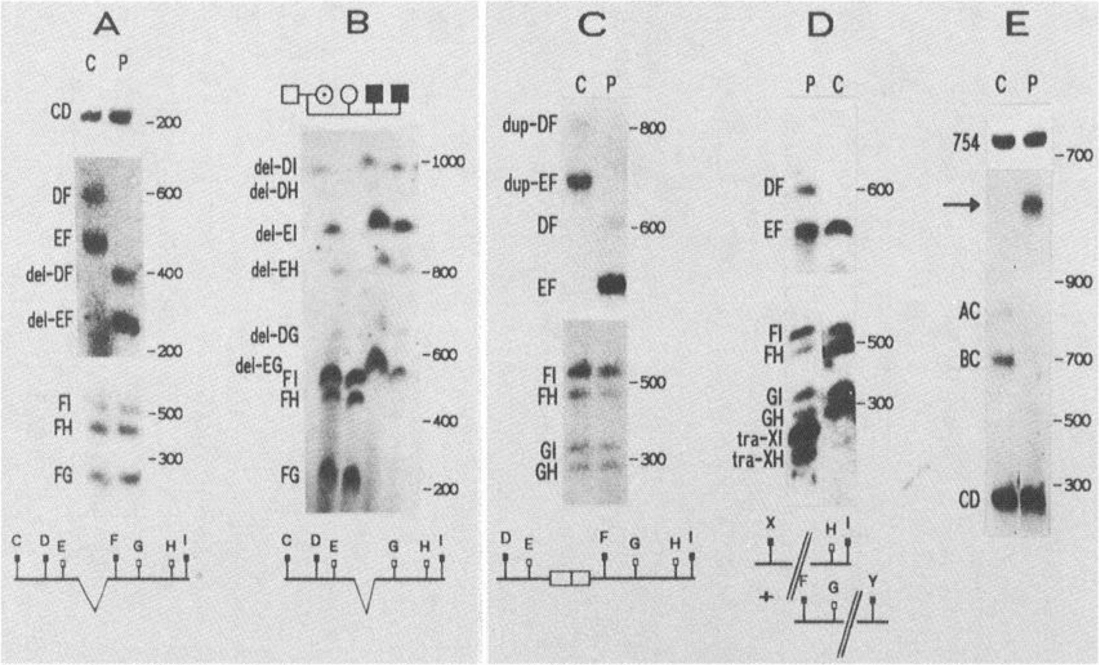


Fig. 3. PFGE analysis of DMD patients and carrier females. DNAs (C = control, P = patient) were digested with SfiI and electrophoresed using standard conditions (Chapter 17). SfiI sites and SfiI fragments are marked with letters (cf Fig. 2). The site of the mutation detected is indicated on the SfiI map below the autoradiographs. (A) FIGE analysis of deletion patient DL23.4. Blots were hybridized with cDNA3b–5a (top), P20 (DXS269, middle), or GMGX11 (DXS239, bottom). (B) CHEF blot of deletion family DA20 (kindly provided by C. van Broeckhoven) hybridized with cDNA5b-7. (C) FIGE blot of duplication patient DL150.5 hybridized with probes JBir (DXS270, top) or J66 (DXS268, bottom). (D) FIGE analysis of DMD-female VSN1 carrying an X;3 chromosomal translocation. Blots were hybridized with P20 (DXS269, top) or J66 (DXS268, bottom). (E) FIGE blot of patient DL185.1, hybridized with 754 (DXS84, top) or cDNA1–2 (bottom). The arrow indicates the abnormal fragment in the patient.

Although DMD is very rare in females, several cases have been described *(19,20)*. Microscopic analysis of metaphase spreads showed the existence of X-chromosomal translocations in most cases *(19)*. We *(2,5)* and others *(3,20)* have used PFGE analysis to locate the translocation breakpoints in the DMD gene of such females. The analysis of Meitinger et al. *(20)*, who studied 11 individual cases, confirmed the extent of the DMD gene. All DMD-causing translocations were located within the DMD gene. Strikingly, none of them showed alterations of the hybridization pattern with the dystrophin cDNA in a conventional Southern-blot analysis. This is explained by the large proportion of intronic sequences of this gene *(see below)*.

An example of a translocation is VSN1, a cell line of a DMD female who carries a balanced X;3 chromosomal translocation. PFGE analysis shows that the translocation in VSN1 disrupts SfiI fragment GH; two additional fragments of 130 and 180 kbp (Fig. 3D), are detected when hybridized to probe J66 (DXS268). The translocation breakpoint of VSN1 could be placed within 80 kbp distal to SfiI site G *(5)*.

The reciprocal translocation junction can be detected with probe GMGX11 (DXS239). It measures over 1 Mbp (not shown), of which more than 700 kbp are thus derived from chromosome 3.

3.3. PFGE Compared with Conventional Analysis

Southern-blot analysis of the PFGE-detectable rearrangements with the dystrophin cDNA confirmed the type of rearrangement derived from PFGE *(7,8)*. Furthermore, it showed that, of the PFGE-detectable aberrations, all but two *(see below)* involved expressed-gene sequences (Table 1). In addition, two duplications of single exons were missed by PFGE analysis, as a result of their small size. This problem can, in principle, be circumvented by increasing the resolution (*see* Section 4).

Two PFGE rearrangements could not be confirmed by a conventional analysis (Table 1). One case is shown in Fig. 3E. In patient DL185.1, SfiI fragments 754 and CD are normal, although SfiI fragment BC is clearly altered. However, a cDNA analysis showed no anomalies *(7,8)*. Hence, there may be involved only regulatory sequences upstream from the gene, or intron sequences, such as those that determine correct splicing. A translocation or inversion cannot be ruled out, but these are rare and should result in more complex changes on a megabase scale (Fig. 1). However, whatever turns out to be the cause of this rearrangement, the detected aberration is a valuable marker of the affected chromosome.

The main advantage of PFGE analysis over conventional blotting is demonstrated by a direct comparison of both methods (Fig. 4). For patient DL43.7,

Table 1
PFGE Analysis of DMD Patients and Carrier Females[a]

SfiI fragment					Fraction	
B-C	C-D	D-F	F-I	I-J	*d*	*d/N*, %
+	+	+	+	+	22	40%[b]
[ ]	+	+	+	+	4	7%[c]
[	]	+	+	+	3	5%
+	[ ]	+	+	+	1	2%
[	–	]	+	+	2	4%
+	[	–	]	+	1	2%
+	+	[ ]	+	+	5	9%
+	+	[	]	+	9	16%
+	+	+	[ ]	+	6	11%[d]
+	+	+	[	–	1	2%
+	+	+	+	[ ]	1	2%
Total PFGE					33	60%[e]

[a]Summary of the results obtained using PFGE analysis Values given *(d/N)* show the number of patients in each category *(d)* after screening *N* patients/carrier females (*N* = 55) Each aberration detected is shown separately: +, normal SfiI fragment (Fig. 2); – deletion, and [ ], altered fragment.

[b]This number includes two single-exon duplications, not visible in PFGE analysis, and two unclear cDNA alterations

[c,d]Including rearrangements that could not be confirmed with Southern analysis (cDNA hybridizations) (c, two rearrangements, d, one rearrangement)

[e]33/55 or 60%.

the situation is straightforward. PFGE analysis detects SfiI AC and BC fragments that are 60 kbp smaller, whereas conventional blotting shows a deletion of three exon-containing fragments. For the carrier female DL43.3, it is difficult to conclude from the conventional blot whether she has one or two copies of the marked fragments. The PFGE-picture, although not perfect, is fully informative; the mutated and normal fragments are clearly discernible. A conventional analysis would even be further complicated in the case of comigrating fragments or of duplications.

As it turns out, the mother, DL43.1, is both a somatic and a germinal mosaic for this deletion mutation *(21)*, transmitting it to two of her children (DL43.3 and DL43.7). The presence of a decreased amount of an abnormal fragment is detectable by PFGE (not shown); the partial reduction in intensity of bands on a conventional blot in DL43.1 is difficult to distinguish from the normal situation (Fig. 4B).

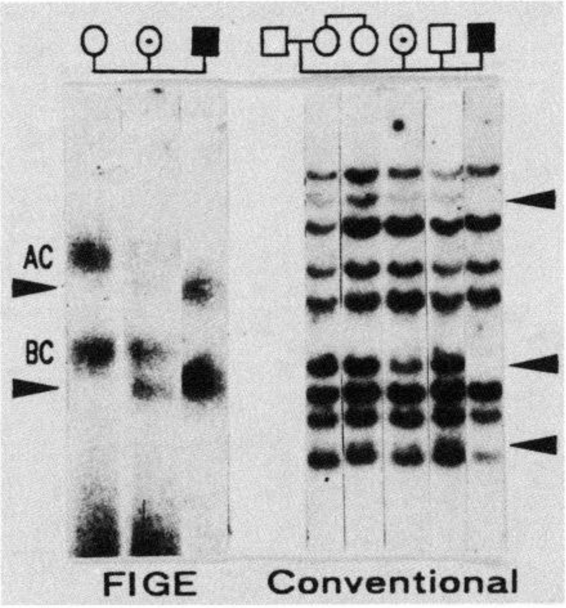


Fig. 4. Comparison of PFGE and conventional Southern-blot analysis of family DL43. The left panel shows a FIGE blot of a SfiI digestion and the right panel a conventional Southern blot of an XmnI digestion hybridized, respectively, with dystrophin cDNA(1–2a) and cDNA(1–2). Arrowheads indicate altered (left) or missing (right) fragments. PFGE bands are marked as in Fig. 2. DNAs used are indicated (top).

3.4. Use of PFGE Data

Reviewing all detected rearrangements makes clear that they are unevenly distributed over the DMD gene (Table 1). The highest fraction of rearrangements map around SfiI site F. A second, but minor, rearrangement hotspot maps around SfiI site C. This prompts the order of probes to use in the analysis; fragments that have a high chance of containing rearrangements are studied first.

Since the PFGE data show the size of a deletion and the cDNA data show which exons are missing, the combination of these data yield a refined map of the exons of the DMD gene (ref. 7, Fig. 2). The exon spacing varies greatly throughout the gene. Intron sizes vary from 107 bp (intron 10; ref. *22*) to some 180 kbp for intron 44 *(7,23)*. In total, the 2.3 Mbp DMD gene consists of 99.4% intron and 0.6% exon sequences.

4. Discussion

In conclusion, PFGE analysis greatly increases the detection "window" around a specific chromosomal location. This potential allows the screening for mutations even without knowledge of the site of the mutation itself. When the borders of any given gene have been put on the physical map and genomic probes are available to detect the fragments involved, one can detect mutations in that gene without the need to isolate the gene. This potential is documented by the analysis of mutations in the DMD gene at a time when the gene itself had not yet been isolated completely *(2–4).*

One limitation of PFGE analysis is in the detection of small rearrangements. The resolution of the analysis can be increased in several ways. First, the time settings of the electrophoresis can be changed to increase resolution in a specific size range. However, lane-to-lane mobility differences, caused by the variation in the amount of DNA loaded per lane *(see below),* do limit resolution. A superior refinement is the study of DNA from carrier females. The presence of the normal fragment as an internal control enables the detection of size differences as small as 5–10 kbp. Practically however, the detection limit of PFGE is about 10–20 kbp.

A second point of caution concerns the mobility of DNA fragments in PFG electrophoresis; both the FIGE (field-inversion) and CHEF (contour-clamped homogenous electric field) variants of PFGE (*see* Chapter 17) are sensitive to overloading with high DNA concentrations, which decrease fragment mobilities. Different DNA concentrations complicate lane-to-lane comparisons and interfere with accurate size estimations.

Third, partial-digestion patterns may complicate the PFGE analysis significantly. The problems raised depend mainly on the quality of the physical map of the region (*see* Section 3.1. and Fig. 3B). Besides inadequate digestion, partial cleavage can be caused by differential sensitivity of the restriction sites. Although it has not been studied in much detail, this problem is probably mainly caused by methylation of nucleotides in the recognition sequence of the restriction enzyme used. The degree of partial digestion, and hence of DNA methylation, has been shown to vary considerably from tissue to tissue *(14).* Analysis of DNA samples from different tissues may decrease this problem. In fact, this analysis has virtues in itself, since it adds valuable new data to the physical map *(14).*

Acknowledgments

We thank E. Bakker and C. van Broeckhoven for their collaboration in the analysis of the DMD patients, P. M. Grootscholten and L. Casula for expert technical assistance, and L. M. Kunkel, M. Ferguson-Smith, and R. G.

Worton for kindly providing probes used in this analysis. We gratefully acknowledge the Dutch Prevention Fund, the Netherlands Scientific Research Organisation, the Muscular Dystrophy Group of Great Britain, and the Muscular Dystrophy Association of America for generous financial support.

References

1. McKusick, V. A. (1989) *Mendelian Inheritance in Man*, 10th, revised Ed., Johns Hopkins University Press, Baltimore, MD.
2. Van Ommen, G. J. B., Verkerk, J. M. H., Hofker, M. H., Monaco, A. P., Kunkel, L. M., Ray, P., Worton, R. G., Wieringa, B., Bakker, E., and Pearson, P. L. (1986) A physical map of 4 million base pairs around the Duchenne muscular dystrophy gene on the human X-chromosome. *Cell* **47**, 499–504.
3. Kenwrick, S., Patterson, M., Speer, A., Fischbeck, K., and Davies, K. E. (1987) Molecular analysis of the Duchenne muscular dystrophy region using pulsed field gel electrophoresis. *Cell* **48**, 351–357.
4. Den Dunnen, J. T., Bakker, E., Klein-Breteler, E. G., Pearson, P. L., and G. J. B. van Ommen. (1987) Direct detection of more than 50% Duchenne muscular dystrophy mutations by field inversion gels. *Nature* **329**, 640–642.
5. Van Ommen, G. J. B., Bertelson, C. E., Ginjaar, H. B., Den Dunnen, J. T., Bakker, E., Chelly, J., Matton, M., Van Essen, A. J., Bartley, J., Kunkel, L. M., and Pearson, P. L. (1987) Long-range genomic map of the Duchenne muscular dystrophy (DMD) gene: Isolation and use of J66 (DX5268), a distal intragenic marker. *Genomics* **1**, 329–336.
6. Chen, J., Denton, M. J., Morgan, G., Pearn, J. H., and Mackinlay, A. G. (1988) The use of Field-Inversion Gel Electrophoresis for deletion detection in Duchenne Muscular Dystrophy. *Am. J. Hum. Genet.* **42**, 777–780.
7. den Dunnen, J. T., Grootscholten, P. M., Bakker, E., Van Broeckhoven, C., Pearson, P. L., and van Ommen, G. J. B. (1989) Topography of the Duchenne Muscular Dystrophy (DMD) gene: FIGE- and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. *Am. J. Hum. Genet.* **45**, 835–847.
8. den Dunnen, J. T., Bakker, E., Van Ommen, G. J. B., and Pearson, P. L. (1989) The DMD gene analysed by field inversion gel electrophoresis. *Br. Med. Bull.* **45**, 644–658.
9. Emery, A. E. H. (1988) Duchenne muscular dystrophy, in *Oxford Monographs on Medical Genetics*, no. 15 (revised Ed.), Oxford University Press, Oxford, England.
10. Monaco, A. P., Neve, R. L., Colletti-Feener, C., Bertelson, C. J., Kurnit, D. M., and Kunkel, L. M. (1986) Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323**, 646–650.
11. Koenig, M., Monaco, A. P., and Kunkel, L. M. (1988) The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell* **53**, 219–228
12. Monaco, A. P. (1989) Dystrophin, the protein product of the Duchenne/Becker muscular dystrophy gene. *Trends Biochem. Sci.* **14**, 412–415.
13. Burmeister, M. and Lehrach, H. (1986) Long-range restriction map around the Duchenne muscular dystrophy gene. *Nature* **324**, 482–485
14. Burmeister, M., Monaco, A. P., Gillard, E. F., van Ommen, G. J. B., Affara, N. A., Ferguson-Smith, M. A., Kunkel, L. M., and Lehrach, H. (1988) A 10 megabase map of human Xp21, including the Duchenne muscular dystrophy gene *Genomics* **2**, 189–202.

15. Robertson, M. (1987) Muscular dystrophy: Mapping the disease phenotype. *Nature* **327,** 372–373.
16. Koenig, M., Hoffman, E P., Bertelson, C. J. Monaco, A. P., Feener, C., and Kunkel, L. M. (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* **50,** 509–517.
17. Forrest, S. M., Cross, G. S , Flint, T., Speer, A., Robson, K. J. H., and Davies, K. E. (1988) Further studies of gene deletions that cause Duchenne and Becker muscular dystrophies. *Genomics* **2,** 109–114.
18. Darras, B. T., Blattner, P., Harper, J. F., Spiro, A. J., Alter, S., and Francke, U. (1988) Intragenic deletions in 21 Duchenne muscular dystrophy (DMD)/ Becker Muscular Dystrophy (BMD) families studied with the dystrophin cDNA: Location of breakpoints on HindIII and BglII exon-containing fragment maps, meiotic and mitotic origin of the mutations. *Am. J. Hum. Genet.* **43,** 620–629.
19. Boyd, Y. and Buckle, V (1986) Cytogenetic heterogeneity of translocations associated with Duchenne muscular dystrophy. *Clin. Genet.* **29,** 108–115.
20. Meitinger, T., Boyd, Y., Anand, R., and Craig, W. (1988) Mapping of Xp21 translocation breakpoints in and around the DMD gene by pulsed field gel electrophoresis. *Genomics* **3,** 315–322.
21 Bakker, E., Veenema, H., den Dunnen, J T., Van Broeckhoven, C., Grootscholten, P M , Bonten, E. J., van Ommen, G. J. B., and Pearson, P. L (1989) Germinal mosaicism increases the recurrence risk for "new" Duchenne muscular dystrophy mutations *J Med Genet.* **26,** 553–559
22. Monaco, A. P., Bertelson, C. J., Liechti-Gallati, S , Moser, H , and Kunkel, L. M. (1988) An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics* **2,** 90–95.
23. Blonden, L A J., den Dunnen, J. T., Van Paassen, H. M. B , Wapenaar M C., Grootscholten P. M., Ginjaar, H. B., Bakker, E., Pearson P. L., and van Ommen G. J B (1989) High resolution deletion breakpoint mapping in the DMD gene by whole cosmid hybridization. *Nucleic Acids Res.* **17,** 5611–5621.

Chapter 27

Molecular Diagnostics of Cancer

Bryan D. Young and Finbarr E. Cotter

1. Introduction

The first consistent chromosomal abnormality to be described in a tumor cell was the Philadelphia chromosome *(1)*, but it was not until the advent of high resolution chromosome banding *(2,3)* that the occurrence of other abnormalities in malignant cells could be fully investigated. Since then, much detailed information has been derived concerning the incidence and nature of molecular abnormalities in human malignancies. Some changes seem to be highly correlated with particular malignancies, whereas others are more general in their incidence. An individual malignancy can often have more than one alteration, with secondary changes superimposed on an original primary event. Many chromosomal translocations, in which genetic material is exchanged between chromosomes, have been documented. Other types of chromosomal alterations can include interstitial deletions, monosomy, trisomy, aneuploidy and the appearance of chromosomes so rearranged as to be unrecognizable. At the submicroscopic level, point mutations in several genes have been documented in tumor cells.

The advent of molecular cloning techniques has provided the possibility of both analyzing these events at the molecular level and exploiting them as unique tumor-specific markers useful in disease management. The small number of changes analyzed so far has led to a better appreciation of the way in which genes important for cell growth can be critically changed as a part of malignant transformation. Among the best-studied examples are the *c-abl* gene in chronic myelogenous leukemia and the *c-myc* gene in Burkitt's lymphoma. These studies support the notion that such consistent changes pinpoint

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

regions of the genome containing genes directly involved in the malignant process. For many documented chromosomal translocations, there are no known suitable candidate genes sufficiently close to the breakpoints. However, recent progress in attempts to map the human genome is providing a multitude of probes, and it can be anticipated that it will be possible to analyze molecularly many chromosome translocations and assess their role in the generation of tumors. The recent development *(4)* of the polymerase chain reaction (PCR) is promising to provide valuable new tools for the diagnosis and monitoring of malignant disease.

2. Methods

The rapid progress in our understanding of the molecular events associated with tumors has depended on the concerted use of a variety of techniques. In order to understand the significance of such findings, it is important to appreciate both the advantages and limitations of each approach. In many instances, the information gained by different approaches is complementary and builds up a total perspective on the course of molecular events.

2.1. Cytogenetic Analysis

The advent of chromosome banding and staining techniques has provided an invaluable means for the recognition of each human chromosome. The identification of a number of specific chromosome abnormalities has led to their molecular analysis, and thus provided the means for a more complete study of these phenomena.

A variety of preparative techniques and staining procedures are commonly used, all with the same objective, viz., the maximal resolution of each chromosome and its subbands. Giemsa banding (G-banding) has become the most widely used technique for the routine staining of mammalian chromosome (*see* Chapter 21). Commonly, slides are treated with a protease such as trypsin *(5)* or hot saline-citrate *(2)*. The resultant chromosome banding patterns are thought to reflect both the structural and functional composition of the chromosomes *(6)*. Quinacrine banding (Q-banding) offers an alternative fluorescence-based approach in which quinacrine dihydrochloride is used as a fluorochrome. The Q-banding pattern of chromosomes appear to be influenced by variations in protein composition *(7)* and is generally similar to that found with G-banding, although there are differences in the centromeric regions of chromosomes 1, 9, and 16 and the acrocentric satellite regions. Staining of the constitutive heterochromatin (C-banding) results in dark staining material in interphase as well as during mitosis. It includes both repetitive DNA, satellite DNA (as detected on centrifugation gradients), and some nonrepetitive DNA. C-banding is said to demonstrate constitutive heterochromatin since satellite DNA has been localized by *in situ* hybridiza-

tion to darkly staining C-band regions *(8)*. For human chromosomes, darkly staining C-bands are located at the centromeres of the chromosomes, with the exception of the Y chromosome, in which the dark band is located on the distal region of the long arm. Marked polymorphism is present in the size of C-bands. Thus, C-banding has application for investigating chromosome rearrangement near centromeres and in investigating polymorphism. Reverse banding (R-banding) results in a staining pattern in which bands that appear pale by G-banding stain darkly by R-banding. Conversely, dark positive G-bands appear pale using R-banding techniques. R-banding can be achieved by incubation in hot saline solution followed by Giemsa staining *(9)*.

Although karyotype analysis performed by any of the above techniques can yield important information about the cancer cell, there are important limitations. A single chromosome band can be reckoned to consist of about 10^7 bp, and therefore, any alteration that involves less DNA than this may be difficult to observe. Additional problems when dealing with tumor tissue can include a low yield of mitoses and poorly banded chromosomes. Usually, a minimum of 20 metaphases will be examined with a higher number being required in difficult cases. A particular advantage of karyotype analysis over Southern analysis *(see below)* is that the whole cell is observed and that multiple events can be documented in a single analysis.

Recently, the technique of nonradioactive hybridization has been developed to the point where it may offer a new form of karyotype analysis based on DNA sequence. By labeling DNA probes with biotin *(10)* or digoxigenin *(11)*, the resultant signal on hybridized chromosomes can be visualized by fluorescence (*see* Chapter 21). Probes can consist of plasmid, phage, or cosmid clones, and although the more repetitive the probe the greater the resultant signal, it is now feasible to detect single copy sequences using probes of only a few kb in length. Probes specific to the repetitive alphoid sequences present at centromeres of chromosomes can be used to examine complex karyotypes *(12)*. A particularly useful application of this technology may be the "painting" of chromosomes by hybridizing with a mixture of probes obtained from chromosome-specific libraries (ref. *13* and Chapter 21, Fig. 5). Potentially, such chromosome-specific "paints" could be prepared for each chromosome and would find application in the analysis of complex karyotypes, such as those found in solid tumors. A further refinement of this approach is the detection of signals in the interface nucleus, thus obviating the need for dividing cells. This has been used for the detection of trisomy 21 *(13)*. In addition to the analysis of chromosomal abnormalities, nonradioactive *in situ* hybridization also has considerable potential for the detection of viral sequences. EBV viral sequence has been detected in nasopharyngeal carcinoma patients *(14)* and cytomegalovirus sequence in biopsies from patients with AIDS *(15)*.

2.2. Southern Analysis of Gene Rearrangements

In Southern analysis (ref. *16* and Chapter 15) genomic DNA is first digested with a restriction enzyme and size-fractionated by electrophoresis on an agarose gel. After transfer to a nylon membrane, the DNA is usually probed with a radiolabeled DNA fragment, thus revealing the pattern of restriction sites on the corresponding genomic DNA sequence. Normally, DNA fragments in the range 10^3 to 20×10^3 bp can be resolved. A reciprocal translocation results in two new junction regions with an abnormal pattern of restriction enzyme sites. Hence, if a DNA probe corresponds to a sequence close to such a junction, Southern analysis of the genomic DNA will reveal an abnormal hybridization pattern. In principle, a suitable DNA probe can be used to detect translocation in tumor cells for which karyotype data is lacking. This approach can only be successful if the breakpoints are known to be clustered within a limited range. Some breakpoints have been shown to occur over a wide range (100 kb), and it would therefore be difficult to use a single probe to detect all such translocations *(17)*. This problem can potentially be solved by using pulsed field gel electrophoresis (*18,19* and Chapter 17) to provide a much larger range of analysis (150–1000 kb). A further problem is that deletions are known to occur around junction regions *(20,21)* and this could result in loss of sequence homologous to the probe and lack of detection of the rearranged allele. Provided such difficulties are taken into account, this approach to detection of translocation can be used to obtain information that cannot readily be acquired by conventional cytogenetics. In particular, a rearrangement can be detected without the need for dividing cells.

Using such approaches, the involvement of certain genes in chromosomal translocations has been well established and these are listed in Table 1. The translocation t(9:22) that generates the Philadelphia chromosome is one of the best studied at both the cytogenetic and molecular level *(22–24)*. The majority of patients with chronic myeloid leukemia (CML) have the characteristic Philadelphia translocation t(9:22)(q34:q11) in their leukemic cells. The oncogene *c-abl*, which is normally present on chromosome 9, is translocated to chromosome 22, where it comes into juxtaposition with the 5' portion of a gene known as the *bcr* or *phl* gene, whose product has an unknown function in normal cells. The molecular consequence of this translocation is the transcription of chimeric mRNA and the expression of a chimeric *bcr–abl* protein with enhanced in vitro tyrosine kinase activity *(23)*. Although the breakpoints on chromosome 9 can occur over a 200-kb range, the breakpoints on chromosome 22 are clustered within a 5-kb sequence rendering this translocation suitable for conventional Southern analysis. In some Philadelphia positive acute lymphoblastic leukemias (ALL), the break in the *bcr* gene has

Table 1
Molecular Analysis of Chromosomal Translocations

Disease	Translocation	Genes	Ref.
CML	t(9;22)	*c-abl/bcr*	*(22,23)*
ALL	t(9;22)	*c-abl/bcr*	*(24)*
Burkitt's lymphoma	t(8;14)	*c-myc*/IGH	*(25)*
	t(2;8)	IGK/*c-myc*	*(26)*
	t(8;22)	*c-myc*/IGL	*(27)*
Follicular lymphoma	t(14;18)	IGH/*bcl-2*	*(28,29)*
B-CLL	t(11;14)	*bcl-1*/IGH	*(30)*
B-CLL	t(14;19)	IGH/*bcl-3*	*(31)*
T-Cell lymphoma	inv(14)	IGH/TCRA	*(32,33)*
	t(7;9)	TCRB/?	*(34)*

The molecular analysis of chromosomal translocations has revealed the involvement of the above genes.

been shown to lie further 5' such that only the first exon of *bcr* is included in the chimeric mRNA *(24)*.

In B-cell leukemias and lymphomas, the immunoglobulin genes have been found to be directly involved in certain chromosomal translocations (Table 1). It was shown that the *c-myc* oncogene was translocated into the heavy chain locus (IGH) *(25)* or, more rarely, into either of the light chain loci in Burkitt's lymphomas *(26,27)*. More recently, the J_H region of the IGH locus has been shown to be involved in the t(14;18) translocation, which is a common feature of follicular lymphoma. This has led to the identification of a gene on chromosome 18 (*bcl-2*), which is directly affected by the translocation *(28,29)*. Similarly, the t(11;14) and the t(14;19) found in B-cell chronic lymphocytic leukemia (CLL) have been shown to involve the IGH locus and molecular analysis has led to the cloning of DNA from the breakpoint of the partner chromosomes *(30,31)*. In T-cell lymphomas an analogous involvement of the T-cell receptor (TCR) genes has been demonstrated for both the chromosomal inversion inv(14) and the translocation t(7;9). The inversion inv(14) is thought *(32,33)* to be a recombination between the TCRA and the IGH loci, whereas the t(7;9) involved a recombination between the TCRB

locus and a region on chromosome 9, which is close to, but does not involve, the oncogene *c-abl* (*34*).

2.3. PCR Analysis

The polymerase chain reaction (PCR) (*4*) has had a dramatic impact on the analysis of genetic disorders of all types. The molecular diagnostics of cancer, in particular, is undergoing a revolution in that relatively simple diagnostic techniques may now be applied to small tissue samples (*35*). In this technique, repeated cycles of specifically primed DNA synthesis are used to amplify the target sequence up to one millionfold or greater. The great fidelity of this reaction means that the product DNA fragments are accurate representations of the original starting sequence. Another key feature of this approach is that the amplification starts only from the sites determined by synthetic oligonucleotide primers chosen by the user. Thus, it is possible to target a short DNA fragment of several hundred base pairs long (from the complete human genome of 3×10^9 bp) and amplify it to almost complete purity. The amplified fragment can then be examined by a variety of means, including a direct reading of its sequence.

PCR has been used for the examination of nucleotide sequence variations (*36,37*), chromosomal rearrangements (*38*), high efficiency cloning of genomic sequences (*39*), direct sequencing of mitochondrial (*40*) and genomic DNA (*41*), and the detection of viral pathogens (*42*).

3. Molecular Diagnostic Applications

3.1. Detection of the t(14;18) Translocation by PCR

The clustering within short regions of the majority of breakpoints on *bcl-2* facilitates the use of PCR for amplification and analysis of t(14;18) breakpoints. By contrast, the variation in the breakpoints around *c-myc* in the t(8;14) make this translocation less suitable for PCR analysis. By positioning oligonucleotide primers on each chromosome adjacent to and oriented toward the expected breakpoint position, it has been possible to amplify specifically only the junction sequences. Since amplification depends on both primers being present on a single DNA fragment, only the recombinant fragment can be amplified. Thus, DNA from normal cells is not able to act as a template for this reaction. *bcl-2* Oligonucleotide primers (either mbr or mcr) flanking the translocation have been used, with a consensus J_H sequence found at the 3' end of each J_H exon, to amplify the 14q+ junctions (*38,43–46*). This approach has been extended to the 18q-junctions using a primer based on part of the recombination signal sequences known to flank germline D_H sequences (*46*). A typical set of oligonucleotide primers is listed in Table 2 for use in amplification of both 14q+ and 18q-junctions in the mbr regions in *bcl-2*.

Table 2
Position, Sequence, and Use of Synthetic Oligonucleotides for Analysis of *t*(14;18) Junctions

Name	Sequence	Use	Position
JH1	5'-ACCTGAGGAGACGGTGACC-3'	PCR	3' J_H consensus
DH1	5'-GTGAGGTCTGTGTCACTGTG-3'	PCR	5' flanking region in D_H
BC1	5'-CCTTTAGAGAGAGTTGCTTTACGT-3'	PCR	5' of mbr in *bcl-2* gene
BC2	5'-ATATTCCATATTCATCACTTTGAG-3'	PCR	3' of mbr in *bcl-2* gene
BC3	5'-CACAGACCCACCCAGAGCCC-3'	SEQUENCING	5' of mbr in *bcl-2* gene
BC4	5'-GTCTGATCATTCTGTTCCCTG-3'	SEQUENCING	3' of mbr in *bcl-2* gene
BC5 (MC12)	5'-GATGGCTTTGCTGAGAGGTAT-3'	PCR	5' of mcr in *bcl-2* gene
BC6	5'-TTATTGAGTGGTCCTTCCTTTC-3'	PCR	3' of mcr in *bcl-2* gene
BC7 (MC7)	5'TCAGTCTCTGGGGAGGAGTGG-3'	SEQUENCING	5' of mcr in *bcl-2* gene
BC8	5'-TCATTTCAGTTGAGTGCTGTG-3'	SEQUENCING	3' of mcr in *bcl-2* gene

Oligonucleotides BC5 and BC7 correspond to MC12 and MC7 used by Ngan et al *(45)* The heptamer recombination signal is underlined in oligonucleotide DH1

Chromosome 18 ←		Chromosome 14 →
Bcl-2	N-region	J1 joining gene
ctccttccgcg	atccggatgtcaaaacccac	aatacttccagcactggggccaaggaac

Fig. 1. A junction between *bcl-2* and the J_1 member of the J_H gene complex determined by PCR and direct sequencing. The position of the N-region is illustrated and gaps have been introduced for clarity.

It has thus been possible to amplify either the 14q+ or 18q- junctions directly from tumor biopsies, marrow samples, or peripheral blood. Since normal cells make no contribution to this reaction, the PCR can be used as a very sensitive test for the presence of cells carrying the t(14;18) translocation. Control experiments indicate that this approach can detect one tumor cell in 10^5 normal cells. The very sensitivity of this assay, however, requires that great care must be taken to avoid contamination with other samples. There is sufficient variability in the *bcl-2* breakpoint, the putative N regions, and the involvement of the J_H gene to render each recombinant fragment essentially unique. Thus, the problem of contamination can be catered for by sequence analysis of the PCR products. This can now be conveniently performed by direct sequencing of the PCR products using the mbr sequencing primers shown in Table 2. This approach avoids the cloning of PCR products and means that a junctional sequence can be read within a few days of receiving a tumor biopsy. A typical result is illustrated in Fig. 1 and demonstrates that in this follicular lymphoma, the *bcl-2* gene has fused through an intervening N-region to the J_1 member of the J_H gene complex. The results for a series of follicular lymphomas with breakpoints in the mbr region (from *[46]*) is illustrated in Fig. 2, and it can be seen that there is considerable variability in the junctional sequences. This analysis and others have indicated that the J_5 and J_6 members of the J_H gene complex are most often involved in these translocations. The variability is useful in that the junctional sequences act as unique clonal markers for each follicular lymphoma. In a recent study *(47)* of patients in long-term remission from follicular lymphoma, this approach was used to examine peripheral blood for the presence of residual lymphoma cells. A proportion of patients with no overt signs of disease were found to have a low percentage of circulating lymphoma cells. Sequence analysis was used to demonstrate that the cells in the peripheral blood were derived from the original tumor mass cryo preserved years previously. The significance of low numbers of cells carrying the t(14;18) translocation in otherwise healthy patients remains uncertain, but has a parallel in the PCR studies of residual cells in patients in remission from B-ALL *(48)*.

Patient A

BCL-2 | J6

CCTTCAGGGTCTTCCTGAAATGCA ctcaccca GATTACTACTACTACTACGGTATGGACGTCTGGGGGCAAGGGACCACGGTCACCGTCTCCTCAGG

Patient B

BCL-2 | J5

CTTCCTGAAATGCAGTGGTGCTTA aaagacgtcc CTAGGTCGACCCCTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGG

Patient D

BCL-2 | J6

GCAGGAAACCTGTGGTATGAAGCC t TGGACATTCTGCCATTGTGATTACTACTACTACTACGGTATGGACGTCTGGGGGCAAGGGACCACGGTCACCGTCTCCTCAGG

Patient E

BCL-2 | DH | J6

CCCAGAGCCCTCCTGCCCTCCTTC ttgagtagatcaaggaggcattacgatattttgactggttattacgctac TACTACTACTACTACGGTATGGACGTCTGGGGCAAAGGGACCACGGTCACCGTCTCCTCAGG

Patient G

BCL-2 | J6

TGGTATGAAGCCAGACCTCCCCGG ggcttagcggggccataaccggca GACGTCTGGGGCCAAGGGACCACGGTCACCGTCTCCTCAGG

Patient L

BCL-2 | J5

CTGTGGTATGAAGCCAGACCTCCC ggg AAGGAACCCTCCTCACCGTCTCCTCAGG

Patient M

BCL-2 | J6

GAAATGCAGTGGTGCTTACGCTCC ggcagcaa ATTACTACTCCTACTACGGTATGGACGTCTGGGGCCAAGGGACCACGGTCACCGTCTCCTCAG

Fig. 2. Sequences of the chromosome 14q+ junctions in seven follicular lymphomas *(46)*. *Bcl-2* sequence is shown in upper case and joining region sequence is in upper case italics with the coding exons in boldface type. Differences between the J_H sequences and their germline equivalents are underlined. The intervening sequences between *bcl-2* and J_H are indicated in lower case. The part of the intervening sequence that is identical to a previously identified D_H region is underlined (patient E).

3.2. Detection of the t(9;22) Translocation by PCR

The positions of the breakpoints in the t(9;22) translocation are less clustered than those of the t(14;18) and, therefore, analysis by PCR requires that the fusion mRNA is first used as a template for cDNA synthesis *(49–52)*. There are three known possible *bcr* junctions with *abl* and, therefore, oligonucleotides for each *bcr–abl* combination have to be designed. It has been variously estimated that one leukemic cell per 10^5 nonleukemic cells *(49)* or 10^6 nonleukemic cells *(52)* can be readily detected by PCR amplification from mRNA. This approach lends itself to the monitoring of leukemic cells following bone marrow transplantation *(51)* or after interferon treatment *(53)*. In both instances, residual leukemic cells were detected in samples from some patients. This approach has also been used to demonstrate that even chronic myeloid leukemia without the translocation expresses the *bcr–abl* fusion transcript *(54)*. The extreme sensitivity of this technique renders it susceptible to the problem of contamination especially in a laboratory where many such reactions are being performed. This problem can be resolved for the t(14;18) translocation by direct sequencing of the PCR products, since each translocation generates unique fusion sequence (Fig. 2). This, however, is not possible for PCR products of the *bcr–abl* fusion amplified from the mRNA and therefore extra care needs to be taken in such experiments.

3.3. Analysis of Clonality by PCR

Although many lymphomas do not have suitable chromosomal translocations on which to perform PCR analysis, about 80% of B-cell malignancies carry only one or two immunoglobulin heavy chain gene rearrangements indicating their clonal origin. The rearrangements of the heavy-chain gene segments during B-cell commitment result in a region called the complementarity-determining region III (CDR-III), which lies between the V_H and J_H regions. This region, which encompasses the diversity region of the heavy-chain segment, because of extensive somatic mutations, provides a DNA-encoded signature specific for each B-cell clone. Suitable V_H and J_H consensus primers flanking this region can be used to amplify by PCR CDR-III sequences from DNA of B-cell population *(55,56)*.

An analogous approach has been developed using the rearrangements to T-cell receptor genes to monitor residual cells in T-cell leukemias *(57,58)*. In a recent study *(48)*, the sequences amplified from leukemias were used to generate diagnostic probes that hybridized only to the amplified CDR-III of leukemic cells from which the sequences were derived. With these probes, leukemic cells could be detected when diluted 1:10,000 with other cells. By cloning the amplified CDR-III into recombinant libraries residual leukemic

cells were accurately quantified in bone-marrow samples from repeated relapses and remissions in one case of acute lymphoblastic leukemia *(55)*. During a clinical remission lasting greater than 7 mo, malignant cells were present in marrow at greater than 1/1000 cells. This approach has been used for accurate quantification of malignant cells in acute lymphoblastic leukemia patients in clinical remission *(48)*, and will allow investigation of the biological significance of low or high numbers of residual leukemic cells in evolution of that disease. In principle, this approach could be used to generate unique clonal markers for any B-cell or T-cell malignancy for which there was no suitable chromosomal translocation.

3.4. Detection of ras *Gene Mutations*

In contrast to the disease specificity of the chromosomal translocations discussed above, about 10% of all human tumors are thought to have acquired mutations to members of the *ras* gene family. These changes have been found to occur at certain positions within the coding sequence, resulting in critical changes to the ras products. The three members of the *ras* gene family, *H-ras*, *K-ras*, and *N-ras* map to chromosomes 11, 12, and 1, respectively. The homologous p21 proteins encoded by this family can bind guanine nucleotides, have intrinsic GTPase activity, and are localized at the inner surface of the plasma membrane *(59)*. They are thought to have a role in the transduction of receptor-mediated external signals into the cell, although the precise biochemical pathway remains to be elucidated. The transforming potential of oncogenic versions of the *ras* genes has been shown to be the result of single-base substitutions that alter the corresponding amino acid and result in reduced GTPase activity *(60,61)*. These point mutations have been found in either codon 12, 13, or 61 of members of the *ras* gene family *(62)* in tumor cells and were not found in normal cells from the same patients. In contrast to some of the specific chromosomal rearrangements discussed above, mutations to *ras* genes have been found in a wide variety of human tumors with varying frequency. One of the highest incidences (25–50%) has been reported in acute myelocytic leukemia (AML) *(62)*.

It is clear that although the majority of mutations in hemopoietic malignancies have occurred in the *N-ras* gene, both *K-ras* and *H-ras* can be affected. Some of the mutations have been found in cell lines and therefore could have arisen in culture. However, there are clear examples of leukemias in which the mutation was present in the primary tumor material. The high frequency of activation of *N-ras* in AML has not been matched by a similar frequency in other myeloid or lymphoid malignancies *(63)*. For example, none of 14 myeloid CML blast crises were found to have mutated *ras* genes *(64)*. It is also apparent that there is no obvious correlation between *ras* muta-

tion and either AML subtype (FAB classification) or karyotypic alteration. It is therefore difficult to establish the role of *ras* mutation in the origin and progression of these tumors. It is of interest that *N-ras* mutations have been demonstrated in three out of eight patients with the myelodysplastic syndrome (MDS) *(65)*. Since it is difficult to predict when MDS will evolve into overt leukemia, it would be important to show whether the presence of a *ras* mutation could predict a leukemic transformation.

3.5. Detection of Tumor Viruses by PCR

The human papilloma viruses (HPV-16 and HPV-18) have been reported to be present at a high frequency in invasive squamous cell cancers of the cervix and in other genital cancers *(66)*. PCR assays have been used to detect levels of HPV sequences at less than one genome per cell *(67)*. This assay can be modified to distinguish between the various HPV subtypes, by first amplifying with consensus primers and then probing with oligonucleotides particular to each subtype. A variant of HPV-16 (designated HPV-16b) has recently been identified by PCR as having a 21-bp deletion *(68)*. Similar PCR based assays have been developed for the detection of other viruses implicated in human cancer. They include the hepatitis B virus (hepatocarcinomas), Epstein-Barr virus, and the human T-cell lymphotropic virus (T-cell lymphomas) *(69–71)*.

3.6. Detection of Gene Amplification

Many types of human cancer have been reported to have amplified copies of particular genes. Two examples in which the amplification appears to correlate with a poor prognosis are the *N-myc* gene in neuroblastoma *(72, 73)* and the *c-erb B-2* gene in breast cancer *(74)*. In addition, the *c-myc* gene has been found to be amplified in cell lines derived from small cell lung cancers *(75)* and in other cancers *(76, 77)*. The epidermal growth factor receptor gene has been found to be amplified in brain tumors of glial origin *(78)*.

In most of these experiments, Southern analysis has been used, with appropriate controls, to quantify gene copy number. The overexpression of a gene can result not only from gene amplification, but from deregulation of an unamplified gene. In this case, the use of Northern blotting with appropriate controls is necessary. PCR technology may offer new possibilities for the detection of both gene amplification and overexpression. A PCR based assay was used to demonstrate the overexpression of thymidylate synthetase mRNA *(79)*. It is possible that similar assays could be developed for overexpression of other genes. Currently, the detection of low levels of gene amplification by PCR remains questionable, owing to the difficulties in performing quantitative PCR assays

3.7. Detection of Minimal Residual Disease (MRD)

Induction therapy in leukemia and lymphoma is administered in order to obtain a complete remission. However, residual disease cells may still remain and have the ability to regenerate a new tumor mass *(80)*. The detection of a marker, such as the Philadelphia chromosome, following allogeneic transplantation for CML does not necessarily herald relapse *(81)* but may be a transient marker representing the presence of residual tumor cells with limited capacity for division as a result of prior treatment. Possibly a failure of an immunological control mechanism or a second promotional event *(82)* may be necessary for further multiplication of the remaining residual tumor cells to occur. Investigation of patients with different stages of lymphoma or leukemia using flow cytofluorometric analysis for κ or λ light chain expression *(81)*, or gene rearrangement analysis suggests that clonal evidence of disease may be found in the absence of clinical or morphological findings *(83,84)*. Studies of the peripheral blood of patients in long-term followup of malignant lymphoma using restriction fragment-length polymorphisms or PCR for the t(14;18) translocation show persistent abnormalities in a proportion despite continuing clinical remission *(38,85,86)*. If the abnormalities are present for many years without the evidence of recurrence, their clinical relevance must be queried.

Quantification of the disease remaining or determination of increased disease bulk at an early subclinical stage may help delineate those patients requiring further therapy caused by early progression while the disease bulk remains small. Those with quiescent disease markers may require no further treatment until evidence of progression is observed *(80,81)*. Southern analysis does allow relative quantification of the percentage of clonal tumor cells present, although the sensitivity is poor in comparison to PCR and may only precede overt clinical relapse by a short period of time *(83)*. Adaptation of PCR to quantification *(87)* will improve the early detection of clonal proliferation of MRD. However, extreme care is essential with the PCR in the detection of MRD, since the sensitivity of the techniques may allow contamination to contribute to false positive results. Direct sequencing of the PCR products will provide unique clonal markers down to the base sequence level for an individual tumor and help reduce the possibility of a false positive result.

Ultimately, defining the gene defect and mechanism of disease may permit the use of therapy directly targeted at the molecular changes. Lymphoid malignancies particularly demonstrating translocations involving Ig or TCR genes, such as the t(8;14) of Burkitt's lymphoma, t(14;18) of follicular lymphoma, and t(8;14) of T-cell neoplasms, where deregulation of gene tran-

scription has been associated with malignant transformation, may be amenable to treatment by "gene therapy." In vitro experiments with antisense oligodeoxynucleotides have successfully demonstrated the ability to decrease *c-myc* protein levels, and thus, malignant proliferation in cell lines containing the abnormal *c-myc* transcripts while leaving the normal *c-myc* protein expression and cell growth unaltered in control normal cells *(88)*. If in vivo trials of this form of "gene therapy" are successful, it will offer great potential for possible curative treatment in an often bad prognostic group of patients, particularly those with MRD destined to relapse. It will become essential to define precisely at a molecular level the disease associated gene rearrangements if this form of treatment is to be considered.

References

1. Nowell, P. C. and Hungerford, D. A (1960) Chromosome studies on normal and leukaemic human leukocytes. *J Natl Cancer Inst.* **25,** 85–88.
2. Sumner, A. T., Evans, H J, and Buckland, K. A. (1971) New technique for distinguishing between human chromosomes. *Nature New Biol.* **232,** 31–32.
3. Patil, S. R., Merrick, S., and Lubs, H. A. (1971) Identification of each chromosome with a modified Giemsa stain. *Science* **173,** 821-823.
4. Saiki, R. K., Gelfand, D. J., Stoffel, S, Scharf, S J, Higuchi, R, Horn, G T., Mullis, K. B, and Erlich, H. A. (1988) Primer-directed amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487-491.
5. Seabright, M. (1972) The use of proteolytic enzymes for the mapping of structural rearrangements in the chromosomes of man. *Chromosoma* **36,** 204–210.
6. Holmquist, G., Gray, M, Porter, T., and Jordan, J. (1982) Characterization of Giemsa dark and light band DNA. *Cell* **31,** 121–129.
7. Sumner, A. T. (1982) The nature and mechanism of chromosome banding. *Cancer Genet. Cytogenet.* **6,** 59–87.
8. Pardue, M. L. and Gall, J. G. (1970) Chromsomal localisation of mouse satellite DNA. *Science* **168,** 1356–1358.
9. Verma, R. S. and Lubs, H. A (1975) A simple R banding technique. *Am. J. Hum. Genet.* **27,** 110–117.
10. Albertson, D. G., Fishpool, R., Sherrington, P., Nacheva, E, and Milstein, C (1988) Sensitive and high resolution in situ hybridization to human chromosomes using biotin labelled probes: Assignment of the human thymocyte CD1 antigen genes to chromosome 1. *EMBO J.* **7,** 2801–2805.
11. Gentilomi, G., Musiani, M., Zerbini, M., Gallinella, G., Gibellini, D., and La, P. M. (1989) A hybrido-immunocytochemical assay for the in situ detection of cytomegalovirus DNA using digoxigenin-labeled probes. *J. Immunol. Methods* **125,** 177–183.
12. van Dekken, H., Pizzolo, J. G, Kelsen, D P., and Melamed, M R. (1990) Targeted cytogenetic analysis of gastric tumors by in situ hybridization with a set of chromosome-specific DNA probes *Cancer* **66,** 491–497.
13. Pinkel, D, Landegent, J, Collins, C., Fuscoe, J., Segraves, R, Lucas, J., and Gray, J (1988) Fluorescence in situ hybridization with human chromosome-specific librar-

ies: detection of trisomy 21 and translocations of chromosome 4. *Proc. Natl. Acad. Sci. USA* **85,** 9138–9142
14 Lung, M L, Chan, K. H., Lam, W. P., Kou, S. K., Choy, D., Chan, C. W., and Ng, M. H. (1989) In situ detection of Epstein-Barr virus markers in nasopharyngeal carcinoma patients. *Oncology* **46,** 310–317.
15. Andersen, C. B. (1990) Detection of cytomegalovirus-infected cells in autopsy material by in situ hybridization. *Apmis* **98,** 363–368.
16. Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98,** 503–517.
17. Leibowitz, D., Schaefer, R. K., Popenoe, D. W., Mears, J. G., and Bank, A. (1985) Variable breakpoints on the Philadelphia chromosome in chronic myelogenous leukemia. *Blood* **66,** 243–245.
18 Schwarz, D. C and Cantor, C. R (1984) Separation of yeast chromosome sized DNA by pulsed field gradient gel electrophoresis. *Cell* **37,** 67–75.
19. Westbrook, C. A., Rubin, C. M., Le Beau, M. M., Kaminer, L. S., Smith, S. D., Rowley, J D., and Diaz, M. O. (1987) Molecular analysis of TCRB and ABL in a t(7;9)-containing cell line (SUP-T3) from a human T-cell leukemia. *Proc. Natl. Acad. Sci USA* **84,** 251–255.
20 de Klein, A., van Agthoven, T., Groffen, C., Heisterkamp, N., Groffen, J., and Grosveld, G. (1986) Molecular analysis of both translocation products of a Philadelphia-positive CML patient. *Nucleic Acids Res.* **14,** 7071–7082.
21. Popenoe, D. W., Schaefer, R. K., Mears, J. G., Bank, A., and Leibowitz, D. (1986) Frequent and extensive deletion during the 9,22 translocation in CML. *Blood* **68,** 1123–1128.
22 Groffen, J, Stephenson, J. R., Heisterkamp, N., de Klein, A., Bartram, C R., and Grosveld, G. (1984) Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. *Cell* **36,** 93–99.
23. Ben-Neriah, Y, Daley, G. Q, Mes, M A., Witte, O. N., and Baltimore, D. (1986) The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. *Science* **233,** 212–214.
24. Hermans, A., Heisterkamp, N., von Linden, M., van Baal, S., Meijer, D., van der Plas, D, Wiedemann, L. M., Groffen, J., Bootsma, D., and Grosveld, G. (1987) Unique fusion of bcr and c-abl genes in Philadelphia chromosome positive acute lymphoblastic leukemia. *Cell* **51,** 33–40.
25 Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., and Leder, P. (1983) The human c-myc oncogene: Structural consequences of translocation into the IgH locus in Burkitt lymphoma. *Cell* **34,** 779–787.
26. Taub, R, Kelly, K., Battey, J., Latt, S., Lenoir, G M., Tantravahi, U., Tu, Z., and Leder, P (1984) A novel alteration in the structure of an activated c-myc gene in a variant t(2,8) Burkitt lymphoma *Cell* **37,** 521–528.
27. Hollis, G. F., Mitchell, K. F., Battey, J., Potter, H., Taub, R., Lenoir, G. M., and Leder, P. (1984) A variant translocation places the lambda immunoglobulin genes 3' to the c-myc oncogene in Burkitt's lymphoma. *Nature* **307,** 752–755.
28. Tsujimoto, Y and Croce, C M. (1986) Analysis of the structure, transcripts, and protein products of bcl-2, the gene involved in human follicular lymphoma. *Proc. Natl. Acad. Sci. USA* **83,** 5214–5218.
29 Tsujimoto, Y., Finger, L. R., Yunis, J, Nowell, P. C., and Croce, C. M. (1984) Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. *Science* **226,** 1097–1097.

30. Tsujimoto, Y., Jaffe, E., Cossman, J., Gorham, J., Nowell, P. C., and Croce, C. M. (1985) Clustering of breakpoints on chromosome 11 in human B-cell neoplasms with the t(11,14) chromosome translocation. *Nature* **315**, 340–343.
31. McKeithan, T. W., Rowley, J. D., Shows, T. B., and Diaz, M. O. (1987) Cloning of the chromosome translocation breakpoint junction of the t(14;19) in chronic lymphocytic leukaemia. *Proc. Natl. Acad. Sci. USA* **84**, 9257–9260.
32. Baer, R., Chen, K. C., Smith, S. D., and Rabbitts, T. H. (1985) The mechanism of chromosome 14 inversion in a human T cell lymphoma. *Cell* **43**, 705–713.
33. Denny, C. H., Hollis, G. F., Hecht, F., Morgan, R., Link, M. L., Smith, S. D., and Kirsch, I. R. (1986) Common mechanism of chromosome inversion in B and T cell tumours: Relevance to lymphoid development. *Science* **234**, 197–200.
34. Reynolds, T. C., Smith, S. D., and Sklar, J. (1987) Analysis of DNA surrounding the breakpoints of chromosomal translocations involving the β T cell receptor gene in human lymphoblastic neoplasms. *Cell* **50**, 107–117.
35. Kawasaki, E. and Erlich, H (1990) Polymerase chain reaction and analysis of cancer cell markers. *J. Natl. Cancer Inst.* **82**, 806–807.
36. Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B., and Ehrlich, H. A. (1986) Analysis of enzymatically amplified β-globin and HLA-DQa DNA with allele-specific oligonucleutide probes. *Nature* **324**, 163–166.
37. Bos, J. L, Fearon, E. R, Hamilton, S. R., Verlaan-de Vries, M., van Boom, J H., van der Eb, A. J., and Vogelstein, B. (1987) Prevalence of ras mutations in human colorectal cancers. *Nature* **327**, 293-297.
38. Lee, M. A., Chang, K. K., Cabanillas, F., Freireich, E. J., Trujillo, J M., and Stass, S. A. (1987) Detection of minimal residual cells carrying the t(14;18) by DNA sequence amplification. *Science* **237**, 175–178.
39. Scharf, S. J, Horn, G. T., and Ehrlich, H. A. (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science* **233**, 1076–1078
40. Wrischnik, L. A., Higuchi, R. G., Stoneking, M., Ehrlich, H. A., Arnheim, N., and Wilson, A. C. (1987) Length mutations in human mitochondrial DNA: Direct sequencing of enzymatically amplified DNA. *Nucleic Acids Res* **15**, 529–542.
41. McMahon, G., Davis, E., and Wogan, G. N. (1987) Characterisation of c-Kl-ras oncogene alleles by direct sequencing of enzymatically amplified DNA from carcinogen-induced tumours. *Proc. Natl. Acad. Sci. USA* **84**, 4974–4978.
42. Kwok, S., Mack, D. H., Mullis, K. B., Poiesz, B., Ehrlich, G., Blair, D., Freidman-Kein, A, and Sninsky, J. J. (1987) Identification of human immunodeficiency virus sequences by using in vitro enzymatic amplification and oligomer cleavage detection. *J. Virol.* **61**, 1690–1694.
43. Stetler, S. M., Raffeld, M., Cohen, P., and Cossman, J. (1988) Detection of occult follicular lymphoma by specific DNA amplification. *Blood* **72**, 1822–1825.
44. Cunningham, D., Hickish, T., Rosin, R. D., Sauven, P, Baron, J. H., Farrell, P. J., and Isaacson, P. (1989) Polymerase chain reaction for detection of dissemination in gastric lymphoma. *Lancet* **1**, 695–697.
45. Ngan, B. Y., Nourse, J., and Cleary, M. L. (1989) Detection of chromosomal translocation t(14;18) within the minor cluster region of bcl-2 by polymerase chain reaction and direct genomic sequencing of the enzymatically amplified DNA in follicular lymphomas. *Blood* **73**, 1759–1762.
46. Cotter, F., Price, C., Zucca, E., and Young, B D. (1990) Direct sequence analysis of the 14q+ and the 18q- junctions in follicular lymphoma. *Blood* **76**, 131–135

47 Price, C. G. A., Meerabux, J., Murtagh, S., Cotter, F. E., Rohatiner, A. Z. S., Young, B D., and Lister, T. A. (1990) The significance of circulating cells carrying the t(14;18) in long remission from follicular lymphoma. *Lancet*, in press.
48 Yamada, M., Wasserman, R., Lange, B., Reichard, B. A., Womer, R. B., and Rovera, G. (1990) Minimal residual disease in childhood B-lineage lymphoblastic leukemia. *N. Engl. J. Med.* **323**, 448–455.
49. Dobrovic, A., Trainor, K. J., and Morley, A. A. (1988) Detection of the molecular abnormality in chronic myeloid leukemia by use of the polymerase chain reaction. *Blood* **72**, 2063–2065.
50. Delfau, M. H., Kerckaert, J. P., Collyn, H. M., Fenaux, P., Lai, J. L., Jouet, J. P., and Grandchamp, B. (1990) Detection of minimal residual disease in chronic myeloid leukemia patients after bone marrow transplantation by polymerase chain reaction *Leukemia* **4**, 1–5.
51 Lange, W., Snyder, D S , Castro, R , Rossi, J. J , and Blume, K. G (1989) Detection by enzymatic amplification of bcr-abl mRNA in peripheral blood and bone marrow cells of patients with chronic myelogenous leukemia. *Blood* **73**, 1735–41.
52. Roth, M S., Antin, J H., Bingham, E. L., and Ginsburg, D. (1989) Detection of Philadelphia chromosome-positive cells by the polymerase chain reaction following bone marrow transplant for chronic myelogenous leukemia. *Blood* **74**, 882–885
53 Lee, M. S., LeMaistre, A., Kantarjian, H. M , Talpaz, M., Freireich, E. J , Trujillo, J. M., and Stass, S. A (1989) Detection of two alternative bcr/abl mRNA junctions and minimal residual disease in Philadelphia chromosome positive chronic myelogenous leukemia by polymerase chain reaction. *Blood* **73**, 2165–2170.
54. van der Plas, D. C., Hermans, A. B., Soekarman, D., Smit, E. M., de, K. A , Smadja, N., Alimena, G., Goudsmit, R , Grosveld, G., and Hagemeijer, A. (1989) Cytogenetic and molecular analysis in Philadelphia negative CML. *Blood* **73**, 1038–1044.
55 Yamada, M., Hudson, S , Tournay, O., Bittenbender, S., Shane, S. S , Lange, B , Tsujimoto, Y., Caton, A. J., and Rovera, G. (1989) Detection of minimal disease in hematopoietic malignancies of the B-cell lineage by using third-complementarity-determining region (CDR-III)-specific probes. *Proc. Natl. Acad. Sci. USA* **86**, 5123–5127.
56 Brisco, M. J , Tan, L. W , Osborn, A. M., and Morley, A. A. (1990) Development of a highly sensitive assay, based on the polymerase chain reaction, for rare B-lymphocyte clones in a polyclonal population. *Br. J. Haematol.* **75**, 163–167
57. Hansen-Hagge, T. E., Yokota, S., and Bartram, C. R (1989) Detection of minimal residual disease in acute lymphoblastic leukemia by in vitro amplification of rearranged T-cell receptor delta chain sequences. *Blood* **74**, 1762–1767.
58 d'Auriol, L., Macintyre, E , Galibert, F., and Sigaux, F. (1989) In vitro amplification of T cell gamma gene rearrangements: A new tool for the assessment of minimal residual disease in acute lymphoblastic leukemias. *Leukemia* **3**, 155–158.
59 Varmus, H. E (1984) Molecular genetics of cellular oncogenes *Annu. Rev Genet* **18**, 553–612
60 McGrath, J. P , Capon, D. J , Goeddel, D V., and Levinson, A. D (1984) Comparative biochemical properties of normal and activated human ras p21 protein. *Nature* **310**, 644–649.
61 Gibbs, J. B , Sigal, I S., Poe, M., and Scolnick, E M. (1984) Intrinsic GTPase activity distinguishes normal and oncogenic ras p21 molecules. *Proc. Natl Acad. Sci USA* **81**, 5704–5708.

62 Bos, J. L., Toksoz, D., Marshall, C. J., Verlaan-de-Vries, M., Veeman, G. H., Van der Eb, A. J., Van Boom, J. H., Janssen, J. W. G., and Steenvoorden, A. C. M. (1985) Amino acid substitutions at codon 13 of the N-ras oncogene in human acute myeloid leukaemia. *Nature* **315**, 726–730.
63 Rodenhuis, S., Bos, J.-L., Slater, R. M., Behrendt, H., van't Veer, M., and Smets, L. A. (1986) Absence of oncogene amplifications and occasional activation of N-ras in lymphoblastic leukemia of childhood. *Blood* **67**, 1698–1704.
64. Janssen, J. W. G., Steenvoorden, A. C. M., Lyons, J., Anger, B., Bohlke, J. U., Bos, J. L., Seliger, H., and Bartram, C. R. (1987) Ras gene mutations in acute and chronic myelocytic leukaemias, chronic myeloproliferative disorders, and myelodysplastic syndromes. *Proc. Natl. Acad. Sci. USA* **84**, 9228–9232.
65. Hirai, H (1987) Oncogenes in hematopoietic malignancies and genetic diagnosis *Nippon Rinsho* **45**, 2864–2871.
66 Pfister, H. (1987) Human papillomaviruses and genital cancer. *Adv. Cancer Res.* **48**, 113–147.
67 Young, L. S., Bevan, I. S., Johnson, M. A., Blomfield, P. I, Bromidge, T., Maitland, N. J., and Woodman, C B. J. (1989) The polymerase chain reaction: A new epidemiological tool for investigating cervical human papillomavirus infection. *Br. Med. Jr.* 298, 14–18.
68. Tidy, J. A., Vousden, K. H., and Farrell, P. J. (1989) Relation between infection with a subtype or HPV 16 and cervical neoplasia. *Lancet* **1**, 1225–1227.
69 Saito, I., Servenius, B., Compton, T., and Fox, R. I. (1989) Detection of Epstein-Barr virus DNA by polymerase chain reaction in blood and tissue biopsies from patients with Sjogren's syndrome. *J Exp Med* **169**, 2191–2198.
70 Duggan, D. B., Ehrlich, G. D., Davey, F. P., Kwok, S., Sninsky, J., Goldberg, J., Baltrucki, L., and Poiesz, B. J. (1988) HTLV-1-induced lymphoma mimicking Hodgkin's disease. Diagnosis by polymerase chain reaction amplification of specific HTLV-1-sequences in tumor DNA. *Blood* **71**, 1027–1032.
71. Kwok, S., Erlich, G. D., Poiesz, B. J., Kalish, R., and Sninsky, J. J (1988) Enzymatic amplification of HTLV-1 viral sequences from peripheral blood mononuclear cells and infected tissues. *Blood* **72**, 1117–1123.
72 Brodeur, G. M., Seeger, R. C., Schwab, M., Varmus, H. E., and Bishop, J. M. (1984) Amplification of N-myc in untreated neuroblastomas correlates with advanced disease stage. *Science* **224**, 1121–1124.
73. Seeger, R. C., Brodeur, G. M., Sather, H., Dalton, A , Siegel, S. E., Wong, W. Y, and Hammond, D. (1985) Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. *N. Engl. J. Med.* **313**, 1111–1116.
74. Slamon, D., Godolphin, W , Jones, L. A., Holt, J. A., Wong, S. G , Keith, D. E., Levin, W J., Stuart, S. G., Udove, J., Ullrich, A., and Press, M. F. (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244**, 707–712.
75 Little, C. D., Nau, M. M., Carney, D. N., Gazder, A. F., and Minna, J. D (1983) Amplification and expression of the c-myc oncogene in human lung cancer cell lines. *Nature* **306**, 194–196.
76 Alitalo, K., Schwab, M , Lin, C. C , Varmus, H. E., and Bishop, J. M. (1983) Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma. *Proc. Natl. Acad. Sci USA* **80**, 1947–1950.

77. Kozbor, D. and Croce, C. M (1984) Amplification of the c-myc oncogene in one of five human breast carcinoma cell lines. *Cancer Res.* **44,** 438–441.
78. Libermann, T. A., Nusbaum, H. R., Razon, N., Kris, R , Lax, I , Soreq, H , Whittle, N , Waterfield, M. D., Ullrich, A , and Schlessinger, J. (1985) Amplification, enhanced expression and possible rearrangement of EGF receptor gene in primary human brain tumours of glial origin. *Nature* **313,** 144–147.
79 Kashani-Sabet, M., Rossi, J J., Lu, Y, Ma, J X., Chen, J , Miyachi, H , and Scanlon, K. J. (1988) Detection of drug resistance in human tumors by in vitro enzymatic amplification *Cancer Res.* **48,** 5775–5778.
80. Monnat, R. J. and Loeb, L A. (1989) Mechanism of neoplastic transformation *Cancer Invest* **1,** 175–183
81 Arthur, C K., Apperley, J. F , and Gou, A P. (1988) Cytogenetic events after BMT for CML in chronic phase. *Blood* **71,** 11791 1 86.
82 Diamond, L., O'Brien, T G., and Baird, W M. (1980) Tumor promoters and the mechanism of tumor promotion *Adv. Cancer Res* **32,** 1–74
83 Brada, M., Mizutani, S , and Molgaard, H. (1987) Circulating lymphoma cells in patients with B and T non-Hodgkin's lymphoma detected by immunoglobulin and T-cell receptor gene rearrangement. *Br. J Cancer* **56,** 147–152
84. Katz, F., Ball, L , Gibbons, B., and Chessells, J. (1989) The use of DNA probes to monitor minimal residual disease in childhood acute lymphoblastic leukaemia *Br. J Cancer* **73,** 173–180
85. Crescenzi, M., Seto, M , Herzig, G P., Weiss, P. D., Griffith, R. C , and Korsmeyer, S. J (1988) Thermostable DNA polymerase chain amplification of t(14,18) chromosome breakpoints and detection of minimal residual disease. *Proc. Natl Acad. Sci USA* **85,** 4869–4873.
86 Fearon, E. R , Burke, P. J , Schiffer, C. A., Zehnbauger, B. A., and Vogelstein, B. (1986) Differentiation of leukemic cells for polymorpholeucocytes in patients with ANLL *N Engl J. Med.* **315,** 15–24
87 Abbot, M. A , Poiesz, B J , Byrne, B C., Kwok, S., Sninsky, J J., and Ehrlich, G. D (1988) Enzymatic gene amplification: Qualitative and quantitative methods for detecting proviral DNA amplified in vitro. *J. Infect. Dis.* **158,** 1158–1169.
88 McManaway, M. E , Neckers, L. M., and Loke, S. L. (1990) Tumour-specific inhibition of lymphoma growth by an antisense oligodeoxynucleotide. *Lancet* **335,** 808–811.

CHAPTER 28

The Detection of Latent Virus Infection by Polymerase Chain Reaction

Norman J. Maitland and Caroline Lynas

1. Introduction

Many standard techniques involving electron microscopy, tissue culture, and protein (antigen) analysis have been developed (reviewed in *1*) for virus diagnosis and typing. Since most viral infections can be assigned to a particular virus type by these means, it is now increasingly important to identify pathogenic variants of each type by more sensitive, discriminating techniques. The main drawbacks of the traditional methods are: the time involved to achieve results; the requirement to find a convalescent patient from a particular infection to provide immune serum, and in many cases, the sheer bulk of infected material that had to be extracted for macromolecular analysis. The advent of nucleic acid based techniques showed much promise, in that the genomes of some viruses could be analyzed to produce what appeared to be an unambiguous "fingerprint," based on restriction endonuclease cleavage patterns of the viral DNA *(2)*. However, these techniques were again limited by tissue culture techniques for the virus, the time required to produce a result, and the expense involved in analysis of single samples. In addition, the techniques were limited in resolution to the number of restriction enzyme cleavage sites that could be assayed at one time. The ultimate fingerprint of any virus would be determination of its complete nucleotide base sequence. This has now been achieved for many viruses, but is a long-term, high cost strategy that could never be implemented in the routine laboratory. Finally, and of most importance, conventional techniques based on tissue culture,

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by. C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

protein, or even nucleic acid analysis were limited to the detection of productive pathogenic infection by the many types of virus. Of more prognostic importance would be the ability to identify and study the many latent infections that are present in human and animal population. These have been compared to a biological time bomb awaiting activation by either immunosuppression of the patient or infection with a second agent. The polymerase chain reaction (PCR), and its capacity to greatly amplify small numbers of specific genes from within a mixture of irrelevant or nonhomologous genes, is the ideal technique to allow the necessary increase in sensitivity and resolution with which to study latent infections. The basic method has been described in detail elsewhere *(3)* and its use for the detection of viruses, is in some ways, much simpler than detection of mammalian gene sequences. The virus genes often show little nucleotide base homology with the genes from their host cells. The methods described below are those in use in the authors' laboratory but can be applied with minor modifications to the detection of many individual types of viruses, often from minute tissue samples taken from apparently uninfected patients.

Detection of viruses by PCR has revolutionized our ideas about the prevalence of latent viral infections, and will undoubtedly increase our ability to analyze productive infections in vivo, without recourse to tissue culture systems. With the added sensitivity comes the need for added precautions, of which only some will be described in detail in this chapter. The potential for generation of false positives is considerably greater in virus detection by PCR, in which the desired levels of detection are frequently less than one genome copy per cell (in contrast to cell gene detection at 1–2 genome copies/cell, for example). If every effort is made to eliminate sources of artefact, by establishing rigorous working practices in the laboratory, then PCR derived results should, in our opinion, truly revolutionize the study of virology.

1.1. Design of PCR Primers for Virus Detection

The design of primers is one of the most crucial factors determining the success of PCR as a method for viral detection. When detecting and analyzing mammalian genes by PCR, you are guaranteed at least one to two copies of the gene per cell. Detection of latent virus infections, on the other hand, often involves amplification from a single virus gene in less than 10,000 mammalian cells. Under these conditions, it is important that the oligonucleotide primers for the PCR either do not self-anneal or anneal to one another. In our experience, primers with more than five base matches will anneal to each

other rather than to the viral gene target. Beyond this it is not possible to lay down more than very general parameters for primer selection, since it is still somewhat empirical. However, the following guidelines will help achieve successful primers:

1. Where possible, select primers with an equal ratio of AT:GC.
2. Avoid primers with stretches of polypurine or polypyrimidines.
3. Check the primers against each other for complementarity, particularly at the 3' end (this can result in primer dimer formation).
4. Primers should be 20–30 bases in length (any random 16–mer should only occur once in the human genome). A larger primer will increase the chance of its being gene-specific. In addition, longer primers can be annealed and extended at higher temperatures, which maintains primer specificity and maximizes *Taq* polymerase activity. If shorter primers must be used, perhaps based on a small stretch of amino-acid sequence, they must be annealed at a lower temperature and similarly, the *Taq* polymerase extension reaction should be carried out at a lower temperature.
5. To compensate for minor sequence variation often found in different virus strains, it is advisable to first choose a gene or part of the gene whose sequence conservation is likely to be very high, or to design primers with a degree of degeneracy within the nucleotide sequence. A number of mismatches within a 20-bp oligonucleotide can be tolerated, although in certain positions, mismatch can result in failure of the PCR.
6. To aid further manipulation or cloning of the amplified fragment, extra sequences that are not complementary to the viral template can be added to the 5' ends of the primers. These exogenous sequences may be amplified as part of the final PCR product and allow the introduction of new restriction enzyme sites, or promoters for RNA polymerase, and so on *(4)*.
7. There do not appear to be any significant differences between primers that will operate on DNA templates and those that will operate on RNA templates. It is possible, however, in those situations where an intron is removed from the mature RNA, to design primers located in exons that will distinguish between the final PCR product produced from DNA and RNA *(5,6)*.
8. Should any particular pair of primers fail to react: if possible, either alter the conditions of the reaction *(see below)* or redesign the primers individually. Frequently, a move of only 10–20 bases 3' or 5' to the original location may be all that is required to achieve success.

2. Materials

2.1. Materials for Sample Preparation

1. Hirt buffer: 0.01*M* Tris-HCl, pH 8, 0.01*M* EDTA, 0.6% sodium dodecyl sulfate (SDS).
2. TE: 10m*M* Tris-HCl, 1 m*M* EDTA, pH 8.
3. GIT: 4*M* Guanadinium isothiocyanate, 50 m*M* Tris-HCl, pH 7.6, 10 m*M* EDTA, 2% w/v sarcosyl, 1% w/v 2-mercaptoethanol.
4. CSTFA: Cesium trifluoroacetate (Pharmacia, LKB) diluted to the required density in TE buffer and 100 µg/mL ethidium bromide.
5. TEN: 10 m*M* Tris-HCl, 1 m*M* EDTA, 0.1*M* sodium chloride, pH 8.

2.2. Materials for RNA PCR

1. 5× Annealing buffer: 100 m*M* Tris-HCl, pH 8.3, 50 m*M* potassium chloride.
2. 10× Reverse transcriptase (RT) buffer: 450 m*M* Tris-HCl, pH 8.3, 375 m*M* potassium chloride, 10 m*M* dithiothreitol, 60 m*M* magnesium chloride, 4 m*M* each of dATP, dCTP, dTTP, and 8 m*M* dGTP.
3. 10× Supplementary buffer: 500 m*M* Tris-HCl, pH 8.3, 166 m*M* ammonium sulfate, 100 m*M* 2-mercaptoethanol, 67 µ*M* EDTA (*see* Note 5).
4. *Taq* polymerase purchased from Cetus (Amplitaq®), Avian myeloblastosis virus reverse transcriptase purchased from Pharmacia/LKB.

2.3. Materials for PCR Sequencing

1. 10× Buffer for sequencing PCR products: 600 m*M* Tris-HCl, pH 7.5, 90 m*M* $MgCl_2$, 100 m*M* dithiothreitol.
2. ^{32}P γ-ATP.
3. Polynucleotide kinase (*see also* Chapter 3 for sequencing reagents).

3. Method

3.1. Sample Preparation

The sensitivity of the PCR *(3)* allows detection of a single viral DNA molecule in a single tissue section or in the quantity of the cells that can be obtained from a single cervical smear. In many cases, it is not necessary to extensively purify nucleic acid prior to carrying out the PCR. Indeed, it has been shown that PCR works directly on tissue culture cells that have been lysed by boiling, and several authors have suggested *(7)* that the method may be extended to clinical samples such as blood and cell smear samples. However, we find that such an approach can lead to unreliable results unless care-

fully controlled (*see* Section 3.1.4). The presence of cell debris becomes inhibitory to the *Taq* polymerase as the number of target cells is increased, but yet a relatively large number of cells may be required to establish a reliable result, when the virus genome of interest is present only at low levels in each cell or in a small proportion of the cells in a tissue sample. Sample preparation is critical to the success and reproducibility of PCR in seeking virus genomes at low levels in a small tissue biopsy. The methods described below have been employed successfully in this laboratory for several different types of tissue samples: (a) fresh biopsy material, (b) blood mononuclear cells, and (c) epithelial cell scrapings. Where fresh tissue samples are difficult to come by, much information of the epidemiology of viruses lies embedded in paraffin tissue blocks. A single section from these blocks can be readily dewaxed and its content of viral genes investigated by the PCR as described in Section 3.1.4. Finally, in some situations, we wish to investigate both the DNA and the RNA contents of the same sample. In this case, a somewhat more sophisticated purification feature is performed, as described in Section 3.1.5.

3.1.1. Tissue Biopsy

1. Slice the frozen material and mince with sterile scalpel blades. Place in an Eppendorf tube containing 1× TE.
2. Boil the sample for 1 h, and centrifuge briefly at 10,000 rpm (room temperature) in a microcentrifuge.
3. Purify the supernatant on a Sephadex column (*see* Section 3.1.6) before PCR analysis.

3.1.2. Blood Mononuclear Cells or Tissue Culture Cells

1. Purify lymphocytes from 5 mL of whole blood using a Ficoll gradient, taking extreme care not to lyse any red cells, since heme compounds are toxic to the PCR *(8)*. The separated cells may be stored as a frozen pellet.
2. Thaw or resuspend the cells in minimum vol of PBS, add 250 µL of Hirt buffer, transfer to a 1.5-mL Eppendorf tube, and mix well.
3. Add 1.5 µL of proteinase K (20 mg/mL in TE), mix well, and incubate at 37°C for 30 min.
4. Add a further 250 µL of Hirt buffer and mix well. Extract 3× with an equal vol of a 1:1 mix of phenol:chloroform.
5. Precipitate the DNA by the addition of 1/10th vol of 5*M* ammonium acetate and an equal vol of isopropanol. Stand at room temperature for 30 min and pellet the DNA by centrifugation at 10,000 rpm for 10 min.
6. Wash the DNA pellets twice in 70% ethanol, dry, and redissolve in TE buffer.

3.1.3. Smears and Scrapings of Epithelium

1. Scrapings of mucosal tissue from either oral or cervical sites can be collected by standard means. In most cases, the yield of cells will be approx 10^5/scrape.
2. The scraping spatula is then rinsed in Tris-buffered saline (in preference to phosphate buffered saline), and the cells centrifuged off on to the base of a conical disposable centrifuge tube. Use of disposable materials should be encouraged at all times, and probably more important, the smear taker should be informed about the risks of cross-contamination. This step is critical in the prevention of cross contaminations, as it is frequently beyond the control of the laboratory scientists.
3. At this stage, the cell pellet can be boiled (*see* Section 3.1.1), after resuspending in either 1x TE or 1x PCR salts to an approximate concentration of 10^6 cells/mL (which should yield about 1 μg of DNA in 25 μL of final suspension). Alternatively, we get more reproducible results by scaling down the proteinase K/Hirt buffer procedure (Section 3.1.1), and adding tRNA carrier (1–5 μg/tube) to precipitate the submicrogram amounts of DNA present in smears that contain very few live cells. This does not interfere with the final PCR reaction.
4. The ultimate nucleic acid pellets, after phenol extraction and isopropanol precipitation (*see* Section 3.1.2) are washed twice in fresh 70% ethanol and redissolved in sterile 1x TE buffer (25–50 μL/sample).

3.1.4. Paraffin Sections

Method A.

1. Place the sections in a 1.5-mL Eppendorf tube and dewax with about 400 μL of xylene. Vortex tube for 1 min and pellet sections by centrifugation for 5 min at 10,000 rpm, decant the xylene and wash the sections three times with absolute ethanol, and allow to dry.
2. For a single 3–5-μm section, proceed as follows: Heat the sections to 95°C for 10 min in up to 40 μL of water followed by the PCR buffer, primers, and enzymes and commence the reaction.

Method B. Depending on the source of tissue, method A may not release sufficient DNA in a form amenable to PCR. A more thorough method of DNA purification is then necessary.

1. Place dewaxed sections in a 1.5-mL Eppendorf tube with 425 μL of TEN, 50 μL proteinase K (20 mg/mL), and 50μL of 10% SDS.
2. Incubate overnight at 55°C.

3. Extract with phenol/chloroform and precipitate the DNA as described in Section 3.1.2.

Method C. Again, some tissue types are reluctant to yield nucleic acid for the PCR by either of the above methods. The protocol below (contributed by James Nichol) has been demonstrated to be particularly efficient for neurological samples, in which fixation times are normally extensive.

1. Transfer sections to clean Eppendorf tube, add 400 µL of xylene, vortex for 1 min, centrifuge for 5 min at 10,000*g*, and pipet off the xylene.
2. Add 400 µL of 100% ethanol, vortex for 1 min, centrifuge as before for 5 min, and pipet off the ethanol.
3. Add 400 µL of 100% ethanol, vortex for 1 min, pipet off the ethanol and vacuum-desiccate the sections for 15 min.
4. Add 50 µL of filtered, double-distilled H_2O, cover with 100 µL of paraffin oil, and incubate for 10 min at 95°C.

Aliquots from this sample can then be used directly in a standard PCR reaction.

3.1.5. Recovery of DNA and RNA Simultaneously from a Single Tissue Biopsy

This method is suitable for all types of samples mentioned so far except paraffin blocks (which are unlikely to contain high yields of useful RNA unless fixed rapidly and with care, as described in ref. *9*), and can deal with biopsies that range in size from several grams to a single mouse ganglion (to which 1–5 µg of carrier tRNA should be added).

1. The sample, snap-frozen in liquid nitrogen immediately on removal from the patient, is homogenized from frozen in 3 mL of GIT buffer.
2. Load the homogenate on to a CSTFA gradient, which comprises 0.65 mL of CSTFA density 1.75g/mL layered on to 0.6 mL of CSTFA density 1.5 g/mL in a 5mL ultracentrifuge tube (treat the ultracentrifuge tube for 30 min prior to this with a 0.2% aqueous solution of diethyl pyrocarbonate to inactivate ribonucleases).
3. Centrifuge the gradients for 16 h at 150,000 × *g* (40,000 rpm, 18°C in a Sorvall AH650 rotor) that will pellet the RNA to the bottom of the tube while banding the DNA at the CSTFA density interface.
4. Using UV light to visualize the nucleic acid, remove the DNA band to a siliconized glass centrifuge tube with a pasteur pipet. Dilute by adding 2 mL of TE buffer and extract three times with phenol/chloroform. Fi-

nally, precipitate the DNA as described in Section 3.1.2. The RNA pellet is finally redissolved in 0.5 mL of 7.5*M* guanidinium hydrochloride, to which 12.5 µL of sterile 1*M* acetic acid and 0.3 mL of ice-cold absolute ethanol are added to reprecipitate the RNA *(10)*.

3.1.6. Sephadex Column Purification of Crudely Extracted DNA (11)

Several potent inhibitors of *Taq* polymerase copurify with DNA. They can be removed by 5 min of boiling followed by passage through a Sephadex column.

1. Prepare a slurry of Sephadex G50 in TEN buffer and centrifuge in a 1-mL syringe plugged with siliconized glass wool at 1600*g* for exactly 3 min. Repeat until the packed Sephadex vol in the syringe is 1 mL.
2. Equilibrate the column with 100 µL aliquots of TEN by centrifuging under exactly the same conditions as above.
3. Add 100 µL of sample to the column and recentrifuge.
4. Elute the DNA from the column into a clean 1.5-mL Eppendorf tube by washing with a further 100 µL aliquot of TEN at 1600*g* for 3 min.

3.2. Detection of Viral DNA by PCR

The precise PCR protocol employed varies according to the enzyme source and the oligonucleotide primers employed. In all cases, the reaction is carried out with between 15 ng and 1 µg of DNA target in a total vol up to 50 µL, containing the quantity of enzyme recommended by the manufacturer. In general, we also use the buffer conditions specified for a particular source of enzyme although we have not found it important to include such components as Tween, NP40, gelatin, and so on. Indeed, in some instances their inclusion has been detrimental to the efficiency of the reaction. For some sets of primers it has been necessary to increase the magnesium chloride concentration of the buffer to as high as 10 m*M*. There is no simple indicator of optimum magnesium concentration, and titration is always necessary for primers that do not work well in the standard buffer (1.5 m*M* magnesium). Indeed, we have two sets of primers that are located 3 kb apart on the genome of herpes simplex virus type 1 DNA. One set works well in standard buffer and the other has a magnesium chloride optimum of 8–10 m*M*.

Standard thermal cycling profiles are 1 min denaturation at 94°C, 1 min annealing at 50°C, and 1 min extension at 72°C, but in practice, these times can be reduced. Overlong time segments result in multiple amplification products from mismatched sequences as the polymerase prefers to amplify something rather than nothing. Although all reaction times can frequently be shortened to as little as 30 s, it is important not to reduce the denaturation

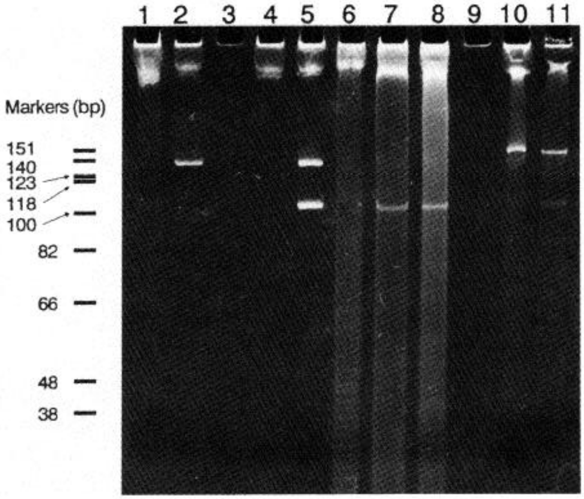


Fig. 1. Detection of Herpes simplex virus DNA and RNA by PCR (data from ref. *5,6*). The figure shows detection of a range of HSVl genes and their mRNA using the standard DNA PCR (tracks 1–5) and RNA PCR (tracks 6–11): analyzed on a 12% polyacrylamide gel. (1) Uninfected cell DNA; (2) HSV1 infected cell DNA + primers for HSV1 ICPO gene (product size 144 bp); (3) as (2) but pretreated with 20 U of DNaseI; (4) HSV1 infected cell DNA with different ICPO primer set (product size is 922 bp and is beyond the analytical range of the PAGE); (5) HSV1 infected cell DNA + primers for both ICPO (144 bp) and thymidine kinase (TK product = 110 bp) genes; (6) uninfected cell RNA + TK primers; (7) HSV1 infected cell RNA + TK primers (110 bp product); (8) as (7) but RNA pretreated with 20 U DNAseI; (9) as (7) but RNA pretreated with 20 µg/mL RNAse A; (10) HSV1 infected cell RNA with same primers as in (4), this time detecting the RNA-specific product, owing to intron removal, of 157 bp; (11) HSV1 infected cell RNA + primers for ICPO (157 bp) and TK (still 110 bp since no intron is removed in mRNA production). Note the relative amounts of ICPO vs TK, which is characteristic of early stages of infection; at later times the TK signal is substantially increased, relative to ICPO.

step to such an extent that the tube contents will not reach 94°C, which is required for complete denaturation of the template. Reduction of cycle time is also not possible with large amplification products, although the polymerase will incorporate about 60 bases/s. Adequate time must be allowed for synthesis, because incomplete amplification products will not be able to serve as template in further cycles. The PCR products are normally visualized on 0.8% agarose or 12% polyacrylamide gels depending on their size (*see* Fig. 1). Up to one-half of the reaction mix is normally loaded. If the resulting bands are faint and the results inconclusive, a 2-µL aliquot of the remainder of the PCR reaction is applied as a template in a second amplification, to confirm the result.

3.3. Detection of Viral RNA by PCR

Since the genomes of many animal viruses are composed of RNA, rather than DNA, the ability to detect this macromolecule by PCR is also important for the virologist. Alternatively, detection of specific classes of mRNA from a viral infection can allow differentiation of a permissive infection (in which late, structural viral genes will be expressed) from a latent infection or a viral transformation (in which only a select subset of principally early viral genes are expressed).

Whereas the *Taq* DNA polymerase is capable of copying mRNA as well as DNA *(12)*, its error proneness often leads to multiple products from the first, critical cDNA synthesis step. Therefore, the modification to the basic DNA PCR protocol described below, which employs reverse transcriptase for the first cDNA synthetic step, was developed to overcome the multiple bands that can complicate analysis of transcription patterns.

3.3.1. RNA Preparation

Any standard method that produces RNA suitable for reverse transcription can be employed as described above (Section 3.1.5). We find that ultracentrifugation through a CSTFA salt gradient removes most, but not all, of the DNA from the preparation, unlike methods that simply harvest cytoplasmic nucleic acid *(13)*. The extreme sensitivity of the PCR means that even a slight contamination (<1pg) with DNA will be sufficient to give a false positive for expression, where, in fact, no viral RNA may be present. This can be overcome by careful choice of primer locations for the PCR (*see* Notes 2 and 3).

Integrity and concentration of the RNA is then assessed by agarose gel electrophoresis before use in the PCR. This step is not absolutely necessary, as the RNA PCR has worked perfectly well with RNA that is largely degraded (e.g., from paraffin blocks, ref. *8*). If necessary, the RNA is finally either reprecipitated or lyophilized again, before either aliquoting or redissolving to a working concentration of 1 µg in 2 µL.

3.3.2. RNA PCR Protocol

1. An annealing mix is first prepared for each reaction to be carried out by mixing 1 µL of 5× annealing buffer and 2 µL of 1 µ*M* primer (*see* Note 2) with 5–10 U of placental ribonuclease inhibitor (e.g., RNAguard, from Pharmacia). The mixture is then incubated at 37°C for 10 min.
2. One microgram aliquots of total cell RNA are then added to the annealing mixes, mixed well and heated to 80°C for 3 min, to fully denature both primer and RNA.
3. Annealing is carried out at 50°C for approx 45 min (*see* Note 4), in the presence of fresh RNA guard (5–10 U) if necessary.

4. To generate cDNA, 2 µL of 10× RT buffer, 5 U of AMV reverse transcriptase (from Pharmacia/LKB, *see* Note 4), and 15 µL of Analar water are added to each annealed mix and incubated at 50°C for a further 45 min.
5. Each reaction mixture is then prepared for the PCR by the addition of 5 µL of 10× PCR supplementary buffer, 2 µL of a 1 µ*M* solution of the second PCR primer, 23 µL of Analar water, and *Taq* polymerase as directed by the manufacturer.
6. The samples are then subjected to the same PCR cycling conditions as those employed for the same primers on DNA, and the products analyzed in precisely the same way.

3.4. Sequencing PCR Products

Whereas the PCR alone is a useful tool in virology, the information can be substantially enhanced by sequencing the amplified DNA product. Sequencing is the best and only unequivocal proof that the amplified PCR product is genuinely derived from the supposed target gene. With the potential for artefact, this is an important consideration. The alternative, DNA hybridization with a probe that lies between, but not including the original PCR primers, is liable to produce false positive (and negative) results, for a variety of reasons. First, if the PCR products are less than 300 bp, a certain amount of variability in binding to nylon or nitrocellulose membranes is possible, leading to false negative results on Southern blotting. Second, spurious PCR products can also react with the hybridization probe, regardless of whether the probe includes the primer sequences, to produce a false positive result.

Related viruses, or strains of the same virus, tend to share short regions of high DNA sequences homolgy, separated by regions of greater diversity. By using primers (or even "best fit" primers) to the former regions *(14,15)* to obtain amplification, followed by direct sequencing, it is possible to map sites of antigenic variation, or gross divergence in a particular genus. The approach could also be employed to isolate as yet uncharacterized viruses, portions of which can be sequenced from their PCR products. The results should be interpreted with caution, however, in the case of "new" viruses. The use of such common primers is also fraught with dangers. Consider the case in particular of coinfection of a sample with two related viruses. If one virus is in vast excess (perhaps where it is in a replicative state), then it will preferentially hybridize to the PCR primers and the overall result will be biased in favor of the replicating virus. It may however be the nonreplicating virus that is responsible for the pathogenesis!

Direct sequencing of the amplification products from an RNA PCR is also the most precise method of determining splice sites present in mRNA.

The inevitable errors and technical difficulties that are present in S1 nuclease or RNase protection assays *(16)* can be largely eliminated, and since multiple spliced forms give rise to different PCR product sizes, each species can be sequenced individually.

3.4.1. Substrate Preparation for Sequencing Reactions

Many published protocols are already available for direct sequencing of PCR products, (*see* Chapter 3), and all will work with products from viral infections. In general, we avoid at all costs subcloning into ml3 phage before sequencing, to eliminate the errors produced by *Taq* polymerase. All "direct" sequences are normally the result of sequencing the products of three independent PCR reactions, which also eliminates the possibility that a putative point mutation arose as a result of a *Taq* polymerase error in the first round of PCR amplification.

There are several important stages that should ensure successful direct sequencing of PCR products of up to 500 bases: First, the product to be sequenced should be homogeneous, i.e., if a mixture of virus types is present, an uninterpretable sequencing gel pattern will be observed. We routinely prepare our fragments by either polyacrylamide or agarose gel electrophoresis, unless the products of a test aliquot are clearly homogeneous. Elution from polyacrylamide is achieved by incubating the gel slice overnight in TE buffer at 37°C, and concentrating the product to 4 µL in 0.1× TE by ultrafiltration using an Amicon filter (Centrikon 30, Amicon Corp.). The final product can then be sequenced directly, or lyophilized to dryness and redissolved in an appropriately low vol of TE. Both the absolute concentration of the eluted PCR product and concentration of salt are important variables that can affect the success of the sequencing. Elution of PCR products from agarose also follows conventional lines, but extreme care should be taken to eliminate all contaminants that will interfere with the sequencing enzymes. Note also that some glass milk-based elution systems *(17)* will not bind small PCR products efficiently. Again, ultrafiltration, using the Millicell (Millipore) products may be the best method for agarose extraction. Protocols are supplied by the manufacturers. Second, to achieve control over the length of sequence, it is also advisable to eliminate as much of the residual dNTP from the PCR product before sequencing. Gel purification, followed by ultrafiltration, also achieves this in most cases.

3.4.2. Sequencing with ^{32}P

1. One of the PCR primers is end-labeled using γ-^{32}P ATP and polynucleotide kinase *(18)*. The mixture is as follows: 1 µL of 1 µ*M* 20 primer, 2 µL 10× buffer, 15.5 µL double-distilled H_2O, 0.5 µL ^{32}P γ-ATP, and 1 µL (2 U) polynucleotide kinase.

2. Ten nanograms of the labeled primer is then annealed with 100–200 ng of PCR product in a final vol of 12 µL of TE and the sequencing reaction carried out (*5,19,* and Chapter 3), using components directly from a Sequenase kit (US Biochemical Corporation). In order to read the sequence within 30 bases of the labeled primer, the chain termination time is reduced to 2 min. It is important to use fresh sequenase enzyme for this reaction, and to avoid over-handling the sequenase stocks.
3. After separating the products on a 12% or 15% polyacrylamide/urea gel, the sequence can be visualized by direct autoradiography of the undried gel. Note that sequences close to the second (unlabeled) PCR primer are frequently difficult to read, in our experience.

3.4.3. Sequencing with ^{35}S

In general, sequencing with ^{35}S is safer and gives more easily interpretable results, over a longer sequence range. However, the requirement to remove urea from the sequencing gel before drying can be tedious. The use of Hydrogel system (Hoefer Ltd.) can overcome this.

1. DNA template (from a PCR reaction) (0.25 pmol) and 20 pmol of chosen PCR rimer (140 ng of a 20-mer) are mixed together in a total vol of 6 µL and annealed after heating to 94°C.
2. Labeling and chain extension steps are then carried out *(20),* although most of the published protocols are viable (Chapter 3).
3. The products are finally separated on a 8% polyacrylamide/urea gel, which is fixed and dried before exposure directly to X-ray film.

3.4.4. Sequencing RNA Templates

Where the yields of PCR product are consistently poor either from a latent viral infection, or when a primer combination is inefficient, it is possible to further amplify the PCR product specifically, by employing a T7 RNA polymerase reaction *(see above).* It is essential in this protocol to maintain all materials under the normal sterile conditions required for RNA work. The products of this amplification can be directly sequenced *(21)* using reverse transcriptase from an end-labeled primer *(4).*

1. RNA template is prepared from purified PCR product in a standard mixture containing: 6 µL of 5x transcription buffer, 5–10 µL of PCR product, 3 µL of 4 m*M* each of ATP, CTP, GTP, UTP, 3 µL of 50 m*M* dithiothreitol, 1 µL RNAgard (Pharmacia), 1 µL of T7 RNA polymerase, made up to a final vol of 30 µL. The mixture is then incubated at 37°C for 1 h. The reaction is terminated by addition of 1 µL of 100 m*M* EDTA.

2. Approximately 20% of this mixture (6 µL) is then mixed with 3 µL of ^{32}P end-labeled primer, (10 ng), prepared as described above in a final vol of 12 µL containing 0.25*M* KCl, 10 m*M* Tris-HCl, pH 8.3.
3. The mixture is then heated to 80°C for 3 min, centrifuged to collect the solution, 1 µL of fresh RNAgard is added, and the primer/template mixture is finally annealed for at least 45 min at 45°C.
4. Prepare four tubes and add 1 µL of 1 m*M* ddATP (tube A), 1 µL of 1 m*M* ddCTP (tube C), 1 µL of 1 m*M* ddGTP (tube G), and 1 µL of 2 m*M* ddTTP (tube T), while in a fifth tube, prepare the following mixture: 13.2 µL of 1× RT buffer, 20 U of AMV reverse transcriptase, 8 µL of primer/template mixture, and 1 µL of RNAgard. Add 5–6 µL of the sequencing mixture to each tube and incubate at 50°C for 45 min.
5. Terminate the reaction by adding 1 µL of Chase solution, and incubate for a further 5 min.
6. Add 2.5 µL of DNA sequencing stop solution (supplied with most sequencing kits).
7. Load sequencing gel.

In many cases, this method of sequencing can overcome premature terminations during the sequencing reaction, caused by high G + C content in the nucleotide sequence, but its main advantage lies in its ability to specifically reamplify one strand of a PCR product.

3.5. Special Features of Viral Genome Detection

3.5.1. Identification of Different Strains of the Same Virus Type

This is of particular importance to the pathologist. Careful selection of primer locations within conserved regions of the viral genome (e.g., those sequences that encode critical structural features of a protein), and spanning polymorphic regions, will produce either (i) a differently sized PCR product, (ii) a product in which a restriction endonuclease site is missing from the original strain, (iii) a product in which a point mutation can be determined by either Sl nuclease mapping, or better still by RNase protection *(16)*, and finally, (iv) the best evidence for sequence variation is obtained by direct DNA sequencing of the PCR product, without cloning *(see below)*. It is always advisable, however,to carry this out on at least two separate pools of PCR reaction products, as in a mixed infection, it is sometimes possible to preferentially amplify one of the two viral types present in the sample.

In a number of instances, we have primers that produce negative results in some samples where the viral gene is clearly present (by Southern blotting for example). Under these conditions, either a mutation within the primer sequence (which would prevent annealing) is possible, or in cases where we

have confirmed the identity of primers and their target sequences, perhaps a change in the state of the viral DNA within a cell has occurred, which could also inhibit the annealing reaction (*see* Section 3.5.4).

3.5.2. Prevention and Detection of Artefact

The major problem with detection of viral genomes in tissue samples by PCR is the possibility of cross-contamination of samples, which will result in false positives. It is important to remember that this application of the PCR makes use of its selectivity, rather than its ability to work with small cell numbers, and that a contamination of 10^{-15}-fold or more with pure viral DNA will result in a false positive result. Common sources of contamination are both known virus infected tissue samples and cloned viral gene probes, which are commonly found in any virus research laboratory *(22)*. We believe that we have overcome this problem by a number of "good working practices," without resort to extremes of containment for our samples *(23)*. Unfortunately, the first possible source of contamination is also the least controlled, i.e., taking of the clinical sample, in which fresh instruments and containers must be employed. Also, the clinicians who are employed in this part of the procedure must be educated in these new requirements. In addition, handling of cloned viral genes is not permitted in the PCR laboratory area.

Second, we normally aliquot the enzyme, primers, and buffer for a large series of identical PCR reactions in a separate laboratory, which leaves only the addition of the sample DNA, minimizing the possibilities of cross-contamination. Water stocks for PCR are kept separate and aliquoted for one working week only and can be treated with UV light *(23)* if necessary.

Third, positive control samples from a diluted viral DNA standard (to monitor the efficiency of each series of reactions) and the inclusion of uninfected cell and/or no DNA (water) negative controls are both essential in every series of experiments. It is never sufficient to assume that because one set of positive or negative control experiments were successful, that subsequent similar experiments will be identical, even with the precaution of aliquoting the enzyme and buffers.

3.5.3. Confirmation of the Presence of Infectious Virus

To simply detect a 200-bp fragment of single virus gene from a complex genome of many kilobase pairs is probably not sufficient to confirm the presence of infectious virus. Equally, to PCR the entire 150 kb of a herpes virus genome would be impractical. One compromise that we have adopted is to amplify and detect portions of both viral early and late genes, which at least confirms that the virus could be capable of replication and encapsidation. These can often be detected simultaneously in the same reaction, by choosing primers separated by different numbers of bases on the genome. To de-

tect viral infection, which is a slightly different thing, we normally look for mature mRNA for the viral late gene products *(see below)*.

3.5.4. Physical State of Viral Genomes in the Infected or Transformed Cell

Considerable importance has been placed on the physical state of a viral genome during its interaction with the host. For example, retroviruses go through an obligatory integration phase (provirus), whereas it has been proposed that human papillomaviruses preferentially integrate into the chromosomes in cells showing malignant changes. To detect these changes in physical state, the classic Southern blotting approach is normally followed *(24)*. The same analysis with PCR is now possible *(25)* using an inverted or "reverse" PCR, in which a restriction endonuclease digestion of the cellular DNA is carried out, the products self-ligated, and a PCR using primers that synthesize divergent new strands, instead of convergent as normal, is employed. The result is synthesis around the self-ligated circle that will run from the virus DNA into the cellular flanking sequences, which can then be sequenced and/or cloned.

4. Notes

1. In the case of heavily keratinized tissues, in which either boiling or proteinase k digestion may not release nucleic acid from the tissues, an alternative first step is to mount the frozen biopsy on a microtome block, and completely section the tissue into 10-µm sections.
2. Primers for cDNA synthesis. A number of reports have suggested the use of either random hexanucleotides, or oligo dT_{15} to prime the synthesis of cDNA *(26)*. In our experience these approaches may be suitable for random cDNA cloning, but for the analysis of specific mRNA from a viral infection, they simply produce extra inexplicable PCR products in addition to that predicted from the viral gene sequence. By priming with a specific 20-mer in the opposite sense to the mRNA, it is also possible to control the reaction by employing the other PCR primer (which is in the mRNA sense) to prime the cDNA synthesis. This is one way of verifying whether it is RNA that is responsible for the final PCR signal *(9)*.
3. Priming across an intron. The best method of distinguishing a PCR signal from a mRNA template from that generated by a DNA template is probably obtained by choosing the PCR primers, such that they span at least one intron in the gene *(9,26)*, *see* Fig. 1, lanes 4 and 7. In this case, the RNA product will differ in size from that obtained with DNA. Alternatively, the samples can be digested with either DNase or RNase (*see*

Fig. 1, lanes 8 and 9) but we find this less satisfactory, again because of the sensitivity of the PCR, which will result in a signal even from degraded material.

4. Choice of reverse transcriptase. AMV reverse transcriptase was preferred to MuLV reverse transcriptase for two reasons. First, it has a higher pH optimum, similar to that for the *Taq* polymerase, resulting in a simple adjustment of conditions from reverse transcription to amplification. Second, the temperature optimum of the AMV enzyme is higher (50°C), which maintains a higher specificity of cDNA synthesis, and reduces the amount of self-priming of the RNA template. The ultimate result is fewer amplification products to confuse the result of the PCR. Manufacturers differ widely in their definition of a unit of AMV reverse transcriptase. The value given is based on our experience with enzyme purchased from Pharmacia/LKB. Most other sources of RT were satisfactory, but required adjustment of the reaction conditions.
5. PCR supplementary buffer. This was devised to adjust the RT buffer toward that recommended for PCR with *Taq* polymerase supplied by Anglian Biotech. Although the buffer conditions suggested by other manufacturers can differ quite considerably, *Taq* polymerase from all other sources (including Amplitaq® from Perkin-Elmer-Cetus) so far tested works perfectly.

References

1. Mahy, B. W. J. (1985) *Virology, A Practical Approach.* IRL, Oxford, UK and Washington DC.
2. Buchman, T. G., Roizman, B., Adams, G., and Stover, B. H (1978) Restriction endonuclease fingerprinting of herpes simplex virus DNA, A novel epidemiological tool applied to a nosocomial outbreak. *J Infect Dis* **138**, 488–498.
3. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Ehrlich, HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.
4. Maitland, N. J., Bromidge, T., Cox, M. F., Crane, I. J., Prime, S. S., and Scully, C. (1989) Detection of human papillomavirus genes in human oral tissue biopsies and cultures by polymerase chain reaction *Br. J. Cancer* **59**, 698–703.
5. Lynas, C., Cook, S. D., Laycock, K. A., Bradfield, J. W. B., and Maitland, N. J. (1989) Detection of latent virus mRNA in tissues using the polymerase chain reaction. *J. Pathol.* **157**, 285–289.
6. Lynas, C., Laycock, K. A., Cook, S. D., Hill, T. J., Blyth, W. A., and Maitland, N. J. (1989) Detection of herpes simplex virus type 1 gene expression in latently and productively infected mouse ganglia using the polymerase chain reaction. *J. Gen. Virol.* **70**, 2345–2355.
7. Saiki, R. K., Bugwan, T. L., Horn, G. T., Mullis, K. B., and Ehrlich, H. A. (1986) Analysis of enzymatically amplified beta-globin and HLA-DQ alpha with allele-spcific oligonucleotide probes. *Nature (London)* **324**, 164–166.

8 Higuchi, R. (1989) Preparation of samples for PCR, in *PCR Technology* (Ehrlich, H. A., ed), Stockton, New York, pp. 31–38.
9 Jackson, D. P., Lewis, F., Wyatt, J I., Dixon, M. F., Robertson, D., Millward-Sadler, H., and Quirke, P. (1989) Detection of measles virus RNA in paraffin-embedded tissue using reverse transcriptasepolymerase chain reaction. *Lancet* (i), 1391.
10. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J , and Rutter, W. J (1979) Isolation of biologically active RNA from sources enriched in ribonucleases. *Biochemistry* **18,** 5294–5299.
11. Maniatis, T., Fritsch, E., and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
12. Jones, M. D. and Foulkes, N. S. (1989) Reverse transcription of mRNA by Thermus Aquaticus DNA polymerase. *Nucleic Acids Res.* **17,** 8387.
13. Maitland, N. J., Cox, M. F., Lynas, C., Prime, S. S., Crane, I J., and Scully, C. (1987) Nucleic acid probes in the study of latent viral disease. *J. Oral Pathol.* **16,** 199–211.
14. Snidjers, P. J F., van den Brule, A J. C., Schrijnemakers, H F J , Snow, G., Meijer, C J L. M., and Walboomers, J. M. M. (1990) The use of general primers in the polymerase chain reaction permits the detection of a broad spectrum of human papillomavirus genotypes. *J. Gen Virol.* **71,** 173–181
15. Gregoire, L., Arella, M., Campione-Piccardo, J., and Lancaster, W. D. (1989) Amplification of human papillomavirus DNA sequences by using conserved primers *J Clin Microbiol.* **27,** 2660–2665.
16 Myers, R. M , Larin, Z., and Maniatis, T. (1985) Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA, DNA duplexes. *Science* **230,** 1242–1246.
17 Vogelstein, B. and Gillespie, D (1979) Preparative and analytical purification of DNA from agarose. *Proc Natl. Acad. Sci USA* **76,** 615–619.
18. Chaconas, G. and van de Sande, J. H. (1980) 5'-^{32}P labelling of RNA and DNA restriction fragments. *Methods Enzymol.* **65,** 75–88.
19. Higuchi, R., von Beroldingen, C. H., Sensabaugh, G. F., and Ehrlich, H. A. (1988) DNA typing from single hairs. *Nature* **332,** 543–545.
20. Gibbs, R. A., Nguten, P-N , McBride, L. J., Koepf, S. M., and Caskey, C. T. (1989) Identification of mutations leading to the Lesch-Nyhan syndrome by automated direct sequencing of in vitro amplified DNA. *Proc. Natl. Acad. Sci. USA* **86,** 1919–1923.
21. Stoflet, E. S., Koeberl, D D., Sarkar, G , and Sommer, S. S. (1988) Genomic amplification with transcript sequencing. *Science* **239,** 491–493.
22. Lo, Y. -M. D , Mehal, W. Z., and Fleming, K. A. (1988) False positive results and the polymerase chain reaction. *Lancet* **ii,** 679.
23. Kwok, S and Higuchi, R. (1988) Avoiding false positives with PCR. *Nature* **339,** 237–238.
24. Botchan, M., Topp, W. C , and Sambrook, J. (1976) The arrangement of simian virus 40 sequences in the DNA of transformed cells. *Cell* **9,** 269–287.
25. Silver, J. and Keerikatte, V. (1989) Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. *J. Virol.* **63,** 1924–1928.
26. Chelly, J , Kaplan, J-C , Mire, P., Gautron, S., and Kahn, S A. (1988) Transcription of the dystrophin gene in human muscle and non-muscle tissues. *Nature (London)* **33,** 858–860

CHAPTER 29

Mapping Inherited Diseases by Linkage Analysis

Martin Farrall

1. Introduction

Family studies have provided experimental observations enabling geneticists to recognize many human genetic traits and diseases. Single-gene Mendelian traits are usually deduced by straightforward inspection of the data, but sophisticated statistical methods have had to be developed to analyze phenotypes that have more complex modes of inheritance. An ongoing catalog of these traits has been compiled by Victor McKusick for over 30 years; 4344 traits are listed in the eighth edition (1988) of *Mendelian Inheritance in Man* (*1*). The exponential increase in reporting of new human genetic information has led to this data base being computerized, and it is available in daily updated form for geneticists to interrogate via academic networks.

For many years geneticists have been frustrated by being able both to identify an inherited trait or disease by family studies and to propose that it would be caused by mutation in a single gene, but being unable to investigate the underlying genetic pathology of the disorder. Useful genetic-counseling advice could be offered to patients and relatives in some cases, but there arose few opportunities to offer genetic screening, and presymptomatic or prenatal diagnosis.

The recent explosion in molecular genetic technology has provided the tools to extend the analysis of inherited traits from the segregation pattern to

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

cloning of the gene or genes responsible for the trait. In some cases this has provided useful information for genetic counselors helping their clients, as well as insight into the pathophysiology of and potential therapeutic strategies for these conditions. There has been much interest in localizing single-gene mutations to individual chromosomes by researchers aiming to isolate and clone the gene responsible for specific diseases. Frequently there is sparse information as to the underlying biochemical defect in these diseases, and mapping, cloning, and sequencing these genes is one of the few options open for understanding these conditions. These so-called reverse-genetic strategies have had several successes; for example, the gene mutated in cystic fibrosis has recently been cloned, and mutations within the gene have been identified, allowing direct carrier detection and prenatal diagnosis *(2)*.

The first step in isolating by "reverse genetics" the defective gene that is mutated in an inherited disorder is to localize the disease trait to a specific chromosomal region. Human geneticists have two main methods available to them for mapping these traits; genes may be directly mapped when affected individuals carry chromosomal aberrations that physically pinpoint the mutated gene, or indirectly mapped in genetic linkage studies with multiply affected families.

Greig cephalopolysyndactyly syndrome, a condition affecting limb and craniofacial development in humans, was localized directly by examining the karyotypes of two unrelated patients; each was found to carry a different balanced translocation with a common breakpoint at 7p13. For many, perhaps most, inherited traits, no evidence of chromosomal rearrangement is found, and genetic linkage studies provide the sole means of chromosomal localization. This is a practical approach when data from sufficient families with multiply affected members can be collected to provide the raw material for linkage studies, namely informative meioses. This requirement usually limits the linkage approach to relatively common conditions. In contrast, those rare conditions associated with chromosomal translocations have the potential to be profitably analyzed with material from only one patient.

The linkage approach has been successful in mapping many genetic diseases, including heritable cancers (e.g., familial polyposis coli *[3]*, neurofibromatosis *[4,5]*, multiple endocrine neoplasia type 2a *[6,7]*), neuromuscular diseases (e.g., Duchenne muscular dystrophy *[8]*), degenerative neurological disorders (e.g., Huntington's disease *[9]* and Friedreich's ataxia *[10]*), adult polycystic kidney disease *(11)*, and the respiratory and gastrointestinal tract disorder cystic fibrosis *(12–14)*. The next sections discuss the methodology that has been followed in mapping such genes in humans using linkage studies, including both the resources that are necessary in collating the data and statistical topics relevant to the analysis of these data.

2. A Brief History of Genetic Linkage Analysis

2.1. Sweet-Peas and Fruit Flies

The classic work by Bateson et al. in 1905 provided evidence that Mendelian characteristics (petal color and pollen grain shape in the sweet-pea) did not always segregate independently of each other, since they observed an excess of parental gametic combinations over reassociations *(15)*. The inferred exchange of genetic material between chromosomes caused the authors consternation when they considered the implications in terms of the chromosomal theory of inheritance, which was initiated in 1903 when Sutton proposed that genes were carried on chromosomes. De Vries had anticipated these exchanges of genetic material *(16)*; Morgan and Cattell in 1912 interpreted recombination in terms of "crossing over" between homologous chromosomes *(17)*. Sturtevant in 1913 produced a genetic map of several sex-linked loci in *Drosophila*, using the recombination fraction as a measure of physical separation *(18)*. This early work has provided the core methodology, which has been followed subsequently by geneticists constructing linkage maps for a diverse range of species including humans.

2.2. Humans

The first genetic linkage in humans was reported in 1937 by Bell and Haldane, who found linkage between X-linked color blindness and hemophilia *(19)*. Mohr reported the first autosomal linkage between Lutheran and Lewis blood groups in 1954 *(20)*. It is pertinent to note that, in his original analysis, Mohr failed to detect the linkage between these blood groups and myotonic dystrophy in the original family, which was evident when likelihood methods were used in the analysis. Linkage analysis in humans really blossomed only with the discovery of the abundance of DNA polymorphisms coupled with simple experimental means to detect and follow them as they segregated through families.

3. DNA Polymorphisms

For the majority of human DNA (possibly as much as 99%) there is no known function. Mutations that accumulate within this "noncoding" or "anonymous" DNA appear to be, in evolutionary terms, selectively neutral; several classes of DNA polymorphisms have been identified within this DNA, and all segregate as codominant markers.

3.1. Restriction Fragment Length Polymorphisms (RFLP)

The simplest of these polymorphisms is a single base change, which has been estimated to occur randomly about once every 150 bp in noncoding, "anonymous" DNA. These point mutations arise by a variety of mechanisms, but the CpG dinucleotide is particularly susceptible to modification. The cytosine in a CpG dinucleotide is liable to methylation outside of HpaII tiny fragment (HTF) islands, and the methylated derivative is frequently converted to TpG. The base substitution may alter a restriction endonuclease recognition site that contains a CpG (e.g., TaqI or MspI, which recognize TCGA and CCGG, respectively); thus, probing a Southern blot made with appropriately digested genomic DNA will reveal a restriction fragment length polymorphism (RFLP). The CpG "hotspot" for point mutation, coupled with the preferential use of CpG restriction enzymes by researchers searching for RFLPs explains their enrichment in published lists of RFLPs.

Probes detecting RFLPs have been reported for all chromosomes, although several investigators have commented that the X chromosome carries fewer and less informative polymorphisms than do autosomes. RFLPs that detect a base substitution have two alleles, which obviously limits the upper boundary for the level of heterozygosity and informativity with a single polymorphism. However, data from multiple tightly linked markers may be combined into a haplotype, which may well be more informative (unless the alleles detected by tightly linked markers are in linkage disequilibrium, *see* Section 4).

3.2. Hypervariable DNA Polymorphisms

Alec Jeffreys and coworkers at the University of Leicester have identified a novel set of DNA sequences containing short, simple, repetitive motifs *(21)*. Using a "minisatellite core" sequence isolated from the human myoglobin gene to probe genomic DNA blots, many restriction fragments are resolved. The complex and highly polymorphic pattern of DNA fragments constitutes a "DNA fingerprint," which has proved useful in paternity, immigration, and forensic cases (*see* Chapters 22 and 23, this volume).

Each fragment corresponds to an individual "minisatellite" sequence that is dispersed throughout the genome, the "core" sequence being repeated tandemly within each "minisatellite." Several types of minisatellites have been identified, which show sequence similarity in their core sequences. These may be crossover "hotspots" analogous to the Chi sequence that initiates recombination in phage lambda. The mechanisms leading to such frequent variation in the number of tandem repeats of the core sequence is unknown. It is unlikely to be generated by unequal exchange during recombination,

but may be generated by slippage during DNA replication *(22)*. The single-copy sequences flanking several minisatellites have been localized on several chromosomes by *in situ* hybridization, and they are clustered at the telomeres.

Jeffreys' DNA "fingerprints" are potentially useful in linkage studies, since multiple loci may be analyzed simultaneously; typically 30 or more loci may be resolved on a single blot. Uitterlinden et al. have increased the data yield from a single blot by a factor of 10 by resolving fragments in two dimensions using denaturing gel electrophoresis *(23)*. Both systems share an analytical limitation, since fragments corresponding to both alleles at a locus are not usually identified, and alleles detected by the same locus in different families cannot usually be matched (a direct result of the high degree of polymorphism). This unfortunately results in data being "private" to each individual family, so data may not easily be pooled across unrelated families.

These problems have been overcome, since individual hypervariable probes may be cloned by probing a genomic library with a core sequence. The core sequence plus the unique flanking sequence may then be used as a "single-copy" probe, detecting alleles at a single locus. These variable-number-of-tandem-repeat (VNTR) probes are as technically straightforward to use as a conventional RFLP, since the banding pattern is simple, consisting of one or two fragments per individual. VNTR probes frequently reveal a high degree of heterozygosity (commonly >80%), and may be physically localized by standard methods. Data may also be pooled between unrelated families, since all alleles detected by a single probe map to the same locus.

The variation found with "minisatellite" DNA has prompted a search for variation within other repetitive DNA families. The simple sequence $(CA)_n$ is very widely dispersed in mammalian genomes, and shows variation between individuals in the number of CA repeats (i.e., alleles are found with $[CA]_n$, $[CA]_{n+1}$, $[CA]_{n+2}$, and so on, ref. *24*). These polymorphisms are typically analyzed by using the polymerase chain reaction (PCR) to amplify a short (about 250 bp) sequence encompassing the $(CA)_n$ repeat and separating the allelic fragments on a denaturing polyacrylamide gel. The fragments can be detected by autoradiography if radioactively labeled primers are used in the PCR. Several other families of polymorphic simple repetitive sequences have been reported (e.g., $[TTA]_n$, ref. *25*, and Alu variable poly[A], ref. *26*), which are widely dispersed throughout the genome (including the X chromosome), show a high degree of heterozygosity, and are proving to be very useful for linkage studies.

4. Linkage Disequilibrium

Alleles detected by probes that map genetically and physically close to each other are occasionally associated with each other as a direct consequence

of their tight genetic linkage. This is detected by counting haplotype frequencies and comparing these observed frequencies with the expected frequencies, which are the product of the individual allele frequencies. For example, consider two loci, with alleles *A* and *a* and *B* and *b*. If the frequency of both *A* and *B* is 0.5, then the expected haplotype frequency for *AB* chromosomes is $0.5 \times 0.5 = 0.25$. Observation that the haplotype frequency for *AB* is significantly different from 0.25 would indicate that there is allelic association or disequilibrium between alleles *A* and *B*.

Individual alleles of an RFLP arise infrequently by spontaneous mutation, so alleles at two tightly linked loci will remain "coupled" unless recombination between the probes generates new combinations of alleles (haplotypes) on a chromosome. It should be remembered that several other genetic factors, such as admixture and selective pressure, may act at a population level to create and sustain the level of disequilibrium.

This population genetic phenomenon of linkage disequilibrium is usually found only for polymorphisms separated by no more than a few tens of kilobases and has been used to advantage by geneticists in some types of study. Recombination is the principal mechanism that generates new haplotypes that mark the decay of disequilibrium, and, in general, the stronger the disequilibrium, the smaller the recombination fraction and associated genetic distance between the markers.

Following this argument, attempts have been made to use the degree of disequilibrium as a "metric" and deduce the relative order of tightly linked markers and mutations leading to inherited diseases *(27–29)*. The varied degrees of success of these attempts suggests that other genetic factors (admixture and selection), as well as random drift, confound high-resolution genetic mapping with linkage disequilibria.

Linkage disequilibrium between RFLPs and inherited disease has been put to clinical use in modifying risks of individuals being carriers of the cystic fibrosis mutation. The disequilibrium additionally provides "phase" information, which is useful when calculating risks for prenatal diagnosis. Disequilibrium may be a nuisance, however, when it limits the gain in informative capability from typing multiple polymorphisms in a small physical (and genetic) region.

5. Construction of the Human Genetic Map

Solomon and Bodner and Botstein et al. were among the first to suggest that DNA polymorphisms would be sufficiently common to be used both as informative markers dispersed throughout the human genome and to construct a genome-wide linkage map in humans *(30,31)*. It is estimated that 330 RFLPs spaced evenly at 10-centimorgan (cM) intervals would span the human genome.

5.1. *Progress*

In 1973, the genetic map of the human genome compiled at the first international Human Gene Mapping Workshop (HGMW) in Yale comprised 27 Mendelian markers and 55 in vitro markers, with hemoglobin and MNSs incorrectly assigned to chromosome 2 *(32)*.

Recombinant DNA methods have fueled the explosion in human gene mapping in two major ways. Genes that have been cloned may be directly mapped by hybridization to a panel of somatic-cell hybrids or by *in situ* hybridization. Alternatively, genes are indirectly mapped by genetic linkage to RFLPs. The HGMW reconvened in Yale in June 1989 and reported on a total of 1631 mapped genes, 113 fragile sites, and 3300 DNA segments *(33)*. The ongoing efforts to sequence systematically the entire human genome will build on this framework of genetically mapped genes until a unified gene map for humans is completed.

5.2. *Resources*

Three internationally available resources have played a central role in the overall synthesis of the currently detailed human genetic map.

The Human Gene Mapping Library (HGML, Director Ken K. Kidd, Yale, USA), in close collaboration with the DNA committee of the Human Gene Mapping Workshop (HGMW, President Bob Sparkes), have maintained a catalog of DNA probes, their chromosomal assignments and regional localizations, and any RFLPs identified by these probes. Currently, there are approx 2000 polymorphic DNA markers, and HGML maintains an internationally accepted system for numbering DNA probes (so-called D numbers).

At the HGMWs, which are held in alternate years, committees responsible for one or two chromosomes edit data submitted by investigators and attempt to derive an overall consensus map integrating diseases and probes. The committeepersons often have to arbitrate between diverse sources and quality of data. The reports are published and provide a key reference that (hopefully) summarizes the state-of-the-art map. The HGML has provided the additional resource of a continuously updated computer data base that may be interrogated interactively over academic networks. The HGMW data has formed the core of the HGML data base, but much additional detail is added, including laboratory details of probes, their availability, addresses of investigators, and literature references.

At HGM10.5 (held in Oxford, September 1990), a new genome data base (GDB) was launched. This data base, developed by the Welch Medical Library (Johns Hopkins University, Baltimore, MD), will be constantly edited by the committee chairpersons at HGMW and is intended to be accessed by geneticists throughout the academic world (*see* Appendix to this volume).

The Centre d'Etude du Polymorphisme Humain (CEPH, director Jean Dausset, Paris, France) provides another key resource to help map the human genome. CEPH has collected a mapping panel of 40 human nuclear families (usually with grandparents) with at least nine children. DNAs are distributed to members of a "collaborative group" that have expressed an interest in gene mapping. The investigators agree to type completely the families with RFLPs that they become interested in mapping. It is expected that data be returned to CEPH headquarters for pooling, so that consensus (or "consortium") maps may be deduced; these maps should be more detailed and accurate than those constructed with data from a single group. The RFLP data base is checked for errors (as far as is possible) and distributed to all collaborating groups.

Most groups declare an interest in mapping particular regions of the genome, a concentration that invariably results from the location of a particular inherited disease. For example, the localization of cystic fibrosis to chromosome 7q was the stimulus that has resulted in a highly detailed map being generated for the whole chromosome *(34)*. However, two groups have contributed in a general way to mapping the entire genome, principally with anonymous DNA markers. This has resulted in the publication of "primary" human genome linkage maps. Donis-Keller et al. *(35)* has reported a 403-locus map with linkage groups on all chromosomes and White et al. *(36)* distributed a booklet containing details of 255 loci on 17 chromosomes.

Since these pioneering maps, much detail has been added, and published maps at 5- to 10-cM resolution are available for many genomic regions. Most of these probes are freely available for general mapping purposes, and Collaborative Research Inc. (Bedford, MA, USA) markets the probes that comprise the Donis-Keller genome map (*see* Appendix to this volume).

6. Strategies in Searching for Linkage

6.1. Candidate Gene Approach

For some traits there may be a clue as to the location of the gene under investigation; linkage may then be sought with markers that map to this region, or with candidate genes themselves if they have been cloned. For example, a patient was reported with a partial trisomy of chromosome 5q and schizophrenia, and Sherrington et al. *(37)* reported the linkage of DNA markers that map to chromosome 5q to a putative autosomal schizophrenia locus.

Clues may come from hypotheses generated from comparison of genetic maps across species. For example, porcine stress syndrome (PSS) and malignant hyperthermia (MHS) in humans have many phenotypic similarities, and both are inherited as simple Mendelian traits. PSS was found to be tightly linked to glucose phosphoisomerase (GPI), and GPI maps to chromo-

some 19q in humans. A recent linkage study has shown that MHS and markers that map in the GPI region of humans are linked (GPI itself was uninformative), confirming the claim that these two diseases are caused by mutations within homologous genes *(38)*.

6.2. Genome-Wide Searches for Linkage

For many traits there will be no clues as to which region of the genome to screen first. There have been two broad approaches to the search, each of which have advantages and disadvantages.

6.2.1. Systematic/Sequential Searches

This chromosome-by-chromosome approach has succeeded in several instances; the availability of a preexisting map of markers at 10- to 20-cM intervals enables efficient searching with multipoint linkage analysis. The RFLP map is constructed with a number of "intervals" (each spanning 10–20 cM), so that the disease locus will be flanked by a pair of RFLPs wherever it happens to map. The exclusion component of "interval" mapping is particularly efficient, since intervals that do not contain the disease locus will generate apparent double-recombination events, which are unlikely. Only one or two meioses consistent with double-recombination events are necessary to exclude a 10-cM interval.

Problems may arise in regions where markers are sparse or only moderately informative, since data insufficient to exclude or include linkage to an interval will be collected. It is also difficult to ensure that intervals extend to the telomore, although the recent cloning and characterization of human telomeres may soon resolve this.

In general, investigators will choose markers that individually show the highest degree of informativity, but two-allele RFLPs are still useful, since many have been accurately mapped or can be combined to generate informative haplotypes.

6.2.2. "Shotgun" Method

Another strategy involves picking at random DNA probes that individually reveal a high degree of informativity and testing for linkage. This pairwise approach will generate substantial regions of exclusion around each marker, but it is difficult to monitor overall progress if the markers are not themselves mapped. Huntington's disease and adult polycystic kidney disease were mapped by this method *(9,11)*.

An elegant variant of the "shotgun" approach is to use a "minisatellite" probe to "fingerprint" the family and to test simultaneously multiple marker loci for linkage. Jeffreys and coworkers have succeeded in linking hereditary persistence of fetal hemoglobin (HPFH) to a single minisatellite locus; up to 34 loci dispersed throughout the genome could be tracked in one experi-

ment *(39)*. This method is suitable for analyzing only large families with enough informative meioses to prove linkage in isolation, since data cannot be easily pooled.

Few investigators would plan or admit to a purely random search for linkage; rather, they would probably opt to test those highly informative markers that became available, provided they were mapped and dispersed throughout the genome. This work would probably continue in parallel with more systematic searches.

6.3. An Example: Friedreich's Ataxia

The search to localize the gene for Friedreich's ataxia (FRDA) illustrates the alternative strategies and their interplay during the laborious search for linkage. FRDA is a rare autosomal recessive disorder (incidence of 1 in 50,000 in the United Kingdom) resulting in progressive spinocerebellar degeneration during the second decade. Despite much research, there were few clues to the underlying biochemical defect, no method for presymptomatic diagnosis, and no specific treatment. A "reverse genetic" project to localize, clone, and analyze the gene mutation in FRDA was therefore initiated in 1985 at St. Mary's Hospital Medical School in London.

A total of 20 multiply affected families were ascertained, principally through consultations at neurology clinics, but also through a patient data base held by a charitable organization, the Friedreich's Ataxia Group UK. Initially, sibships with at least three affected members were collected. For a recessive disease, meioses from one sibling are "consumed" to establish phase, so a maximum of four informative meioses may be derived from a "3-affected" family at a cost of DNA-typing five individuals (provided both parents are informative). For a "2-affected" family, a maximum of two informative meioses may be deduced after typing four individuals. The efficiency of data collection as judged by the number of informative meioses deduced for each individual being DNA-typed, is 0.8 for a "3-affected" family and 0.5 for a "2-affected" family. Obviously, larger sibships would yield data more efficiently, but they are rare.

Candidate gene: A portion (20%) of FRDA patients develop clinical diabetes mellitus, and pharmacokinetic studies have shown the insulin receptor (INSR) to be present in normal densities, but with a much-reduced binding affinity for insulin. INSR had been previously cloned and mapped to chromosome 19p. Linkage studies with INSR polymorphisms detected obligate recombination events in several FRDA families *(40)*.

Systematic search: The remainder of chromosome 19 was then systematically excluded from being linked to FRDA *(40)*. This region was chosen to commence the structured exclusion study, since premapped probes were

readily available for much of the rest of chromosome 19. These probes detected two allele RFLPs, but were sufficiently informative to exclude the majority of chromosome 19. Other chromosomal regions that were reported to be covered with a number of appropriately (10–20 cM) spaced markers were also examined in turn.

"Shotgun" search: While the systematic searches continued, several highly informative markers (e.g., HLA) were tested for linkage as they became available. A panel of polymorphic protein and red-cell antigen markers were also tested in the families by researchers who had semiautomated assays established and could analyze the FRDA samples at a relatively low cost. Most of these markers were only moderately informative; one notable exception was the MNS blood group system. A few VNTR probes that were highly informative were also analyzed in the families.

Markers covering 80% of the genome (117 markers) were excluded from linkage before a large positive lod score (*see* Section 7, especially 7.2.3) was finally revealed with a probe mapping to chromosome 9 in 1988 *(10)*. There were several instances in which markers showed maximal lod scores of >2.0, and one instance when a lod score nearly reached 3.0, which is broadly consistent with the theoretical false positive rate of 5%, which corresponds to a lod score threshold of +3.0 (*see* Section 7.2.3).

Before the search for linkage was successfully concluded, some neurologists claimed that the clinical (phenotypic) heterogeneity was likely to be reflected in genetic heterogeneity. This could be either intragenic heterogeneity (a number of different mutations within the same gene) or intergenic heterogeneity (mutations in a number of genes that map to different chromosomal regions). To date all FRDA families have proved to be linked to chromosome 9 markers, which argues against intergenic heterogeneity. The tight and homogeneous linkage has been used to clinical advantage in first-trimester prenatal diagnosis of this condition (*41, see also* Chapter 30, this volume).

7. Statistical Considerations

The cardinal principles of good practice for experimental design apply equally to linkage analysis and to any other type of study that will undergo statistical analysis:

1. Hypotheses should be declared at the outset of the study.
2. Appropriate statistical methods and significance levels should be chosen.
3. The sample size should be adequate to ensure that the study has sufficient power to achieve its objectives.

7.1. Hypotheses

In linkage analyses, the null hypothesis (H_0) states that alleles at the disease locus and the RFLP under examination segregate independently; in other words, the recombination fraction between the two loci is 50%. The alternative hypothesis (H_1) might state that the recombination fraction between the two loci is <50% (e.g., 10%). In practice, most investigators do not wish to be confined by anticipating the recombination fraction, and multiple H_1s corresponding to various recombination fractions are implicitly assumed. The H_1 that fits best is chosen and the rest forgotten.

7.2. Analysis and Thresholds

Likelihood or lod score methods of analysis have been effective in their application to analyze human genetic data efficiently and reliably. Methods developed to analyze for experimental organisms the offspring from "ideal" matings are generally of little practical use in analyzing human data; the phase of alleles at multiple loci is rarely known, and data are often missing for key family members.

7.2.1. Likelihood Calculations

The likelihoods of pedigrees with arbitrary structures, including multiple marriages and consanguineous loops, segregating with markers may be calculated with the aid of computer programs. Analyses by hand or with the aid of tables of lod scores are really of use only in the simplest of cases. Programs that have had widespread application in linkage analysis include Liped *(42)*, Linkage *(43)*, and Mapmaker *(44)*. These programs permit a flexible specification of the underlying mathematical model for the segregation of loci through the families.

The mode of inheritance is defined, both for discontinuous (simple Mendelian) and continuous (quantitative) traits. Loci may be autosomal, sex-linked, or pseudoautosomal. Multiple alleles detected at the same locus may be specified together with their associated frequencies. Penetrance, the conditional probability that an individual with a known genotype expresses a phenotype, may be defined, and multiple penetrance classes are used to correct for "age of onset." Phenocopies, individuals with normal genotypes that appear to be affected by nongenetic causes, may also be allowed for. Haplotype frequencies may be incorporated when markers show linkage disequilibrium. Spontaneous mutation rates may also be specified. The Linkage program additionally has an option that calculates genetic risks.

7.2.2. Sex Differences in Recombination

There is extensive evidence that the recombination fraction between a pair of linked loci varies with the sex of the parent. For example, a review of

linkage data for chromosome 1 loci revealed an overall 2/1 female/male ratio in recombination fractions. This is consistent with data from other species (e.g., mouse and *Drosophila*), in which a relative excess of recombination is found in females (homogametic sex) over males (heterogametic), which defines Haldane's law. There are clear exceptions to this rule, with males showing more recombination than females. The ratio may also vary from chromosome to chromosome and between different regions on the same chromosome. Currently available computer linkage-analysis packages allow full specification of male and female recombination rates.

7.2.3. Statistical Inference

Likelihoods are calculated at several recombination fractions and compared with the "null" likelihood, calculated with the recombination fraction set at 50%. The lod score represents a likelihood-ratio test and is expressed as the $\log_{10}$ likelihood difference, i.e., the $\log_{10}$ likelihood at the "test" recombination fraction minus the "null" $\log_{10}$ likelihood.

By convention, lod scores are calculated and reported at several recombination fractions, namely 0.00, 0.01, 0.05, 0.10, 0.15, 0.20, 0.30, and 0.40. The maximal lod score ($\hat{Z}$) and the corresponding maximal likelihood estimate of the recombination fraction ($\hat{\theta}$) are also recorded.

A lod score of +3.0 expresses odds of 1000/1 supporting linkage, and is the threshold value generally accepted as adequate evidence to prove linkage between loci *(45)*. The "raw" odds ratio of 1000/1 corresponds to a final (posterior) probability of 95% that the two loci are truly linked. This calculation takes into consideration the modest prior chance (which is conventionally taken as 1 in 50) that any two loci chosen at random will be linked. A threshold of –2.0 is conventionally chosen as sufficient evidence to exclude linkage between loci. This represents a highly stringent exclusion threshold with a false negative rate of 0.02% (remember that the prior chance of linkage is a low 2%).

It may seem surprising that the accepted exclusion threshold is so much more stringent than the false positive rate. It should be remembered that positive linkages will almost certainly be followed up, by collecting data from additional families and by adding in data for new polymorphisms. By contrast, excluded regions are discarded, and the investigator will continue the search for linkage elsewhere. The risk of missing linkage and scanning the rest of the genome unnecessarily strongly supports the choosing of a highly stringent exclusion threshold.

It is interesting to review reports of linkage in the literature to see how many substantial positive lod scores turn out to be false positives. One such report just failed to link the cystic fibrosis locus to a DNA marker on chromosome 21 in an extended Amish kindred group (maximal lod score = 2.48).

The same family showed overwhelming evidence of linkage to chromosome 7q markers when they were tested later *(46)*. Another recent example involved a report of linkage between chromosome 11 markers (Harvey ras and insulin) and manic depression, with an original pairwise lod score of 4.08 *(47)*. Reanalysis with new data, namely inclusion of new individuals and two changes in clinical status, markedly reduced the lod score *(48)*. Analysis of an additional branch of the family led to a final exclusion of this region of chromosome 11.

This revelation has prompted the editorial staff of *Nature* to speculate whether linkages between loci should be published only if lod scores are >6.0 *(49)*. The author would personally favor a threshold of 3.7 (which represents a 1% false positive rate) to be adopted for analysis of simple Mendelian traits. More stringent thresholds are necessary when analyzing traits that present diagnostic difficulties and uncertainties about penetrance or age of onset. In all studies, investigators should try to collect and analyze data from as many families and polymorphisms as possible in an attempt to publish scores that exceed the threshold comfortably, rather than "give up" when the score just exceeds 3.0.

7.2.4. Multiple Testing

During the search for linkage, a substantial amount of exclusion data will be collected (unless the investigator is extremely lucky). In statistical terms, this represents multiple tests for linkage; each test (i.e., Is the maximal lod score for this pair of loci greater than +3.0?) is associated with a false positive rate of 5%. Thus, after 20 independent tests for linkage, a false positive result is to be expected! This problem of correcting a "primary" significance level to compensate for multiple tests arises in many statistical fields and has been addressed by Ott in relation to linkage *(50)*.

However, two other factors may be considered that compensate (at least partially) for the reduced significance level associated with repeated testing. First, as regions of the genome are excluded, the remaining genome to be scanned is shrinking, and the prior probability of linkage correspondingly increases. For example, if 50% of the genome is excluded, then the prior probability of two loci being linked is 1/25; a lod score of 3.0 (1000/1 odds supporting linkage) is therefore associated with a false positive rate of 25/1000 = 2.5%.

Second, tests for linkage with multiple loci on the same chromosome are statistically interdependent. One test may therefore encompass multiple markers vs the disease locus, so the total number of statistical tests is considerably smaller than the number of markers.

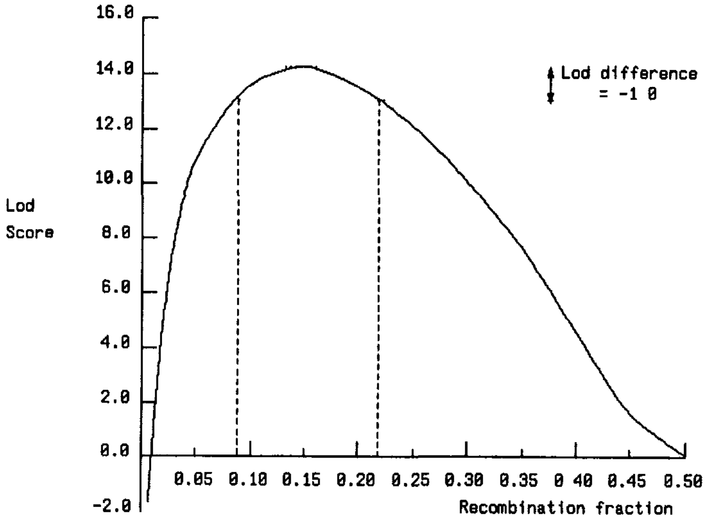


Fig. 1. Lod-score graph illustrating "lod – 1.0 support" method for deducing confidence limits for recombination fractions

7.2.5. Estimation of Recombination

This is conventionally taken as the maximal likelihood estimate (MLE) of the recombination fraction, i.e., the recombination fraction that yields the largest lod score. This may be approximated either by quadratic interpolation or numerically, using an iterative algorithm. The latter method is implemented in the ILINK program from the Linkage package by the Gemini routine (*see* Chapter 31 for a discussion of linkage software).

7.2.6. Confidence Limits

It is useful to express the confidence that investigators should associate with the MLE of a recombination fraction, since estimates may depart from true values with sampling error. This can be done following large-sample theory, but the applicability to typical human data is unclear.

An empiric but simple method that claims to provide a confidence limit of approx 95% is demonstrated in Fig. 1, which shows an illustrative lod score graph for two loci. The MLE of the recombination fraction is 15% with a lod score of 14.15. A line is drawn one lod unit below the maximal score (13.15), lines are dropped perpendicularly from the two points at which this line cuts the likelihood curve, and two recombination fractions are read (8.5 and 21.5%). This "lod – 1.0 support" method follows a convention proposed by

the HGMW in Helsinki *(51)* and is gaining acceptance by the scientific community, through frequent application. The original recommendation was that these limits approximated a 95% confidence limit for "large" samples. The author interprets this as applying to tables generated with more than 30 informative meioses.

7.3. Power of the Study

Studies should be designed so that they have a very good chance of detecting linkage when loci are truly linked. A lod score of 3.0 will theoretically be found between 5% of pairs of unlinked loci. If a lod score of 3.0 is found for a study that has only a 5% chance of reaching a significant score, then the chances of a true positive and a false positive are equal. It is prudent to attempt only linkage studies that have at least a 95% chance of detecting linkage (lod score of at least 3.0) when the loci are truly linked. This may present problems for investigating rare diseases for which only a few families are known.

Each phase-known meiosis contributes a lod score of 0.301 ($\log_{10} 2$) when the loci cosegregate; hence, 10 phase-known meioses are the minimum necessary to attain a lod score >3.0, assuming fully informative markers. For many studies, family structure and mode of inheritance preclude direct deduction of phase. Reduced penetrance, correction for age of onset, and missing data further confound attempts to deduce the effective number of informative meioses (ENIM) in the family.

The ENIM may be estimated quickly and simply by the investigator before any family members are sampled or typed with markers. The pedigree structure is drawn, typings are "invented" for an imaginary, totally informative, highly polymorphic marker that cosegregates infallibly with the disease, and only those members that are likely to be available for sampling are "typed." These data may then be entered into a conventional computer linkage-analysis package for calculation of lod scores, and allowance may be made as appropriate for reduced penetrance, age of onset, phenocopies, and the like. The maximal lod score should be found at zero recombination. This lod is divided by 0.301 to yield the ENIM.

An example of the utility of calculating the ENIM is shown with reference to Fig. 2. Here a pedigree with dominant spinocerebellar ataxia is shown. In this condition, heterozygotes develop symptoms as they grow older, so an age-of-onset correction is necessary. Heterozygotes in each of the four generations have a 100, 90, 75, and 50% chance, respectively, of expressing the "affected" phenotype. The "simulated" genotypings of a highly informative four-allele RFLP are also shown. The maximal lod score (at zero recombination) is 2.06, and the ENIM is therefore 2.06/0.301 = 6.84. Obviously, data from other families would have to be collected before it would be worthwhile initiating a genome-wide search for linkage. Formal power calculations may

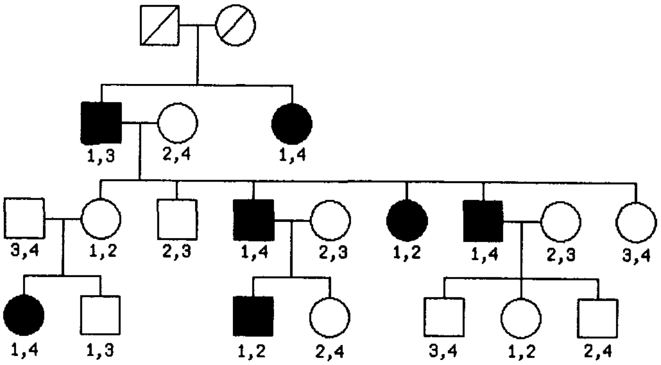


Fig. 2. Spinocerebellar ataxia (SCA) pedigree segregating with a highly informative marker. The dominant SCA gene segregates with the marker allele 1.

be made analytically, but are practical only for simple pedigrees *(50)*. Boehnke has written a computer program for estimating the power of families to detect linkage by repeatedly simulating the family and possible genotypings *(52)*.

8. Heterogeneity

Mutations in different genes may result in very similar phenotypes, and linkage studies have the potential to reveal this genetic heterogeneity. For example, Morton discovered significant linkage heterogeneity between elliptocytosis and the rhesus blood group in 14 families *(53)*.

Likelihood-ratio tests have been devised to test if multiple families are linked to a single locus, and lod scores can be added together. Ott distributes a set of computer programs (HOMOG) that implement these methods *(50)*.

9. Multipoint Linkage Analysis

When family data are available for three or more loci on a chromosome, then attempts may be made to deduce genetic order. For three loci *A*, *B*, and *C* in a line, three recombination fractions (θ_{AB}, θ_{BC}, and θ_{AC}) may be estimated. If these raw recombination fractions are transformed into genetic distances *(d)* using a mapping function, then $dAC = dAB + dBC$. It is simple to deduce the genetic order, provided the estimates of the three recombination fractions are accurate and derived from independent samples of chromosomes. However, multipoint crosses can provide more information for deducing order than the pairwise recombination fractions.

9.1. Multiple Crossing Over

Geneticists working with three-point crosses in experimental organisms noticed that, as a consequence of multiple crossing over, recombination in adjacent intervals was not additive. For example, for the loci *A–B–C*, sequential cross-overs in intervals *AB* and *BC* will be counted in the estimation of θ_{AB} and θ_{BC}, but not in that of θ_{AC}. Double cross-overs in small intervals are uncommon, and the most probable order for a set of loci will show the fewest multiple cross-overs.

9.2. Interference

In crosses in experimental organisms, double cross-overs have been observed less frequently than expected if cross-overs occurred independently of each other. It seems that one cross-over inhibits a second cross-over in the immediate vicinity. This positive genetic interference has been observed in many organisms, including *Drosophila* and mice, and thus is anticipated to occur in humans.

9.3. Mapping Functions

These define a mathematical relationship between recombination and genetic distance (or density of crossing over). They make empiric assumptions as to the frequency of multiple cross-overs, which in turn makes assumptions about the degree of interference. Genetic distances are measured in morgans, 1 cM being equivalent to 1% recombination. This equality becomes inaccurate for recombination fractions greater than about 15%.

9.4. Joint-Likelihood Multipoint Linkage Analysis

The lod score method, which has been used successfully with pairwise linkage data, has been extended to analyze data segregating simultaneously for multiple loci. Lathrop has developed the Linkage program for joint-likelihood analysis of an arbitrary number of loci. In many problems, a single marker is not sufficiently informative to "track" all the meioses in a family; however, data from flanking loci may be analyzed jointly and yield more information overall. This efficient extraction of mapping information from the expensive (in terms of time, labor, and money) data allows more accurate mapping and, frequently, more confidence in interpreting the results. Exclusion of a disease locus from a map of linked markers is particularly efficient, since double cross-overs will be inferred when the disease is located incorrectly (*see also* Section 6.2.1).

In the current version of Linkage, likelihoods for four or more loci are calculated assuming no interference. This has been criticized on the theo-

retical grounds that mathematical modeling with interference would be biologically more accurate and estimates of genetic distances without interference would be exaggerated. In practice, this assumption probably makes little difference. For example, maps constructed by multipoint analysis tend to be slightly larger than those deduced from pairwise data. For investigators attempting to map new loci, the assumption of no interference will minimize the contribution of double cross-overs and make claims of exclusion conservative.

There is no elegant way to tabulate multipoint likelihoods as conveniently as lod scores for pairwise data, in such a way that new data can simply be added in. Usually recalculation with the original pedigree structure and genotypings will be necessary to integrate new data. The support for linkage of a new marker to a preexisting map of marker loci is often graphically expressed as a location map.

One problem faced by all geneticists using joint-likelihood methods for multipoint analysis is the substantial consumption of computer time and memory. Families with genetic diseases frequently have individuals with missing data, who are essential to include since they link informative branches of the family together. Likelihoods have to be calculated for all possible joint genotypes for these individuals. As the number of loci under examination increases, the number of possible joint genotypes increases dramatically. At the present time, the author uses a UNIX workstation with a fast (12-MIPS) RISC processor. There have been many problems that have not been analyzed completely, since they would involve an impractical length of processor time. In these situations, subsets of loci are analyzed jointly and the overall map constructed somewhat empirically from these fragments.

10. Family Collection

The most important and frequently limiting component of a linkage study is ascertaining and collecting families suitable for detecting linkage.

10.1. Autosomal Dominant

Typically, multigeneration families with several affected individuals are sampled. For example, in Huntington's disease, a single large Venezuelan pedigree was collected with sufficient affected individuals for a powerful study. Dominant disorders occasionally show incomplete or age-dependent penetrance, so individuals may carry the mutant allele, but appear phenotypically unaffected. This is a feature of Huntington's disease; carriers develop symptoms only in the fourth decade. This reduces the information contribution of younger family members. For common dominant traits, occasional homozygous affected individuals may well be sampled.

10.2. Autosomal Recessive

Nuclear families are most typically collected for recessive traits. Grandparents are unaffected and, in the absence of a biochemical carrier test, cannot contribute any phase information for the disease. They may be useful for deducing the phase for markers in a multipoint analysis. Pseudodominant families are reported only infrequently, and it should be remembered that homozygotes are uninformative for linkage. Consanguineous matings classically bring together recessive alleles and may be usefully collected. These families provide "phase-known meioses," which are unusual in human genetic-linkage analyzes. Pedigrees with many inbreeding "loops" present analytic difficulties since each loop dramatically increases the calculation time with currently available algorithms.

10.3. X-Linked Traits

Males are hemizygous, which eases deduction of phase.

11. Mapping of Complex Traits

Linkage studies have a proven track record in mapping loci that have a well-defined mode of inheritance. Major genes that cause common disorders, such as the low-density-lipoprotein receptor (LDLR) and familial hypercholesterolemia, have been analyzed in families with a dominant, single-gene mode of inheritance.

There are several common conditions of clinical importance that show familial clustering, but do not show an obvious or consistent inheritance pattern (e.g., atherosclerosis, hypertension, diabetes, cancer, and mental illness). This is probably a consequence of an individual's phenotype being modified by multiple genes (polygenic) as well as nongenetic (environmental) factors.

Methods for statistical analysis that extend the lod score method to map the underlying genes for such traits have been developed, but it is unclear if they will be of practical use with typical human data sets. It seems prudent to attempt to map these complex traits in experimental organisms, for which much larger and controlled data sets can be made available, and then investigate candidate genes or genetic regions in humans.

An alternative analytic approach to searching for linkage to genes involved in complex traits involves affected-relative pair methods. These "identity-by-state" extensions to the "classic" sib-pair method of linkage analysis provide alternative means and strategies for attempts to identify genes that contribute to complex multilocus diseases *(54,55)*.

12. Concluding Remarks

Recombinant-DNA technology has provided abundant polymorphic markers that are suitable for genetic-linkage studies in humans. Statistical methods have been developed that can efficiently analyze the data, so scans of the genome are practical for locating disease loci. Presymptomatic and prenatal diagnosis and carrier detection are feasible for mapped diseases (*see* Chapter 30). Linkage can be used to test for genetic heterogeneity between families. Finally, reverse genetic strategies may then be devised to isolate and clone the underlying gene (*see* Chapters 18 and 19). Understanding the genetic pathology of a disease is the first step in both development of specific therapies and offering prospects for population-based genetic screening.

References

1 McKusick, V. A. (1988) *Mendelian Inheritance in Man*, 8th Ed., Johns Hopkins University Press, Baltimore, MD

2 Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plasvic, N, Chou, J.-L., Drumm, M. L., Iannuzzi, M. C., Collins, F. S., and Tsui, L.-C. (1989) Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA. *Science* **245,** 1066–1073 (*see also* related papers in *Science* **245,** 1059–1065 and 1073–1080).

3. Bodmer, W. F., Bailey, C. J, Bodmer, J., Bussey, H. J. R., Ellis, A., Gorman, P., Lucibello, F. C, Murday, V. A., Rider, S. H, Scambler, P. J., Sheer, D., Solomon, E., and Spurr, N. K. (1987) Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* **328,** 614–616.

4 Barker, D., Wright, E., Nguyen, K., Cannon, L., Fain, P., Goldgar, D., Bishop, D. T., Carey, J., Baty, B., Kivlin, J, Willard, H., Waye, J. S., Greig, G., Leinwand, L., Nakamura, Y., O'Connell, P., Leppert, M., Lalouel, J.-M., White, R., and Skolnick, M. (1987) Gene for von Recklinghausen neurofibromatosis is in the pericentromeric region of chromosome 17. *Science* **236,** 1100–1102.

5 Seizinger, B. R., Rouleau, G. A., Ozelius, L J, Lane, A. H., Faryniarz, A. G., Chao, M. V., Huson, S., Korf, B. R., Parry, D. M, Pericak-Vance, M. A., Collins, F. S., Hobbs, W. J., Falcone, B. G., Iannazzi, J. A., Roy, J. C., St. George-Hyslop, P. H., Tanzi, R. E., Bothwell, M. A., Upadhyaya, M., Harper, P., Goldstein, A. E., Hoover, D. L., Bader, J L, Spence, M. A., Mulvihill, J. J., Aylsworth, A. S., Vance, J. M., Rossenwasser, G. O. D., Gaskell, P. C., Roses, A. D., Martuza, R. L., Breakfield, X. O, and Gusella, J. F. (1987) Genetic linkage of von Recklinhausen neurofibromatosis to the nerve growth factor receptor gene. *Cell* **49,** 589–594.

6 Mathew, C G P., Chin, K. S, Easton, D. F., Thorpe, K., Carter, C., Liou, G. I., Fong, S-L., Bridges, C. D. B., Haak, H., Kruseman, A. C. N., Schifter, S., Hansen, H. H, Telenius, H., Telenius-Berg, M, and Ponder, B A. J. (1987) A linked genetic marker for multiple endocrine neoplasia type 2a on chromosome 10. *Nature* **328,** 527,528.

7. Simpson, N. E., Kidd, K. K., Goodfellow, P. J., McDermid, H., Myers, S., Kidd, J. R., Jackson, C. E., Duncan, A. M V., Farrer, L A., Brasch, K., Castiglione, C., Genel, M., Gertner, J., Greenberg, C R., Gusella, J. F., Holden, J. J. A., and White, B. N. (1987)

Assignment of multiple endocrine neoplasia type 2a to chromosome 10 by linkage. *Nature* **328**, 528–530.
8 Davies, K. E., Pearson, P. L., Harper, P. S., Murray, J. M., O'Brien, T., Sarafazi, M., and Williamson, R. (1983) Linkage analysis of two cloned DNA sequences flanking the Duchenne muscular dystrophy locus. *Nucleic Acids Res.* **11**, 2303.
9 Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottima, K., Wallace, M. R., Sakaguchi, A. Y, Young, A B , Shoulson , I , Bonilla, E , and Martin, J. B. (1982) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234–238.
10. Chamberlain, S , Shaw, J , Rowland, A., Wallis, J., South, S , Nakamura, Y , von Gabain, A., Farrall, M., and Williamson, R. (1988) Mapping of mutation causing Friedreich's ataxia to human chromosome 9. *Nature* **334**, 248–250.
11 Reeders, S T., Breuning, M. H., Davies, K. E., Nicholls, R. D., Jarman, A. P., Higgs, D., Pearson, P. L., and Weatherall, D. J. (1985) A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16 *Nature* **317**, 542–544.
12 Tsui, L.-C., Buchwald, M., Barker, D., Braman, J C., Knowlton, R , Schumm, J W , Eiberg, H , Mohr, J , Kennedy, D., Plasvic, N., Zsiga, M , Markiewicz, D , Akots, G , Brown, V., Helms, C., Gravius, T., Parker, C , Rediker, K., and Donis-Keller, H. (1985) Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science* **230**, 1054–1057.
13 Wainwright, B J., Scambler, P J , Schmidke, J , Watson, E A , Law, H -Y, Farrall, M., Cooke, H. J., Eiberg, H., and Williamson, R. (1985) Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* **318**, 384–386.
14 White, R., Woodward, S , Leppert, M , O'Connell, P , Nakamura, Y, Hoff, M , Herbst, J , Lalouel, J.-M., Dean, M., and Vande Woude, G. (1985) A closely linked genetic marker for cystic fibrosis. *Nature* **318**, 382–384.
15. Bateson, W., Saunders, E. R., and Punnett, R. C. (1905) Experimental studies in the physiology of heredity. *Rep. Evol. Comm. R. Soc* **2**, 1–55, 80–99
16. De Vries, H. (1910) Fertilization and hybridization, in *Intracellular Pangenesis,* C. S. Gager, Chicago, pp. 217–263
17 Morgan, T. H., and Cattell, E. (1912) Data for the study of sex-linked inheritence in *Drosophila J. Exp Zool* **13**, 79–101
18 Sturtevant, A. H. (1913) The linear arrangement of six sex-linked factors in *Drosophila,* as shown by their mode of association. *J. Exp Zool* **14**, 43–59
19 Bell, J. and Haldane, J. B S. (1937) The linkage between the genes for colour-blindness and haemophilia in man. *Proc R. Soc.* **B123**, 119–150, and reprinted in *Ann Hum. Genet.* **50**, 3–34 (1986).
20. Mohr, J. (1954) *A Study of Linkage in Man.* Munksgaard, Copenhagen.
21. Jeffreys, A., Wilson, V , and Thein, S. (1985) Hypervariable "minisatellite" regions in human DNA *Nature* **314**, 67–73
22 Jeffreys, A. J., Neumann, R , and Wilson, V. (1990) Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis *Cell* **60**, 473–485
23. Uitterlinden, A. G., Slagboom, P. E., Knook, D. L., and Vijg, J. (1989) Two dimensional DNA fingerprinting of human individuals. *Proc. Natl. Acad. Sci. USA* **86**, 2742–2746.
24 Litt, M and Luty, J. A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac actin gene. *Am J Hum Genet* 44, 397–401

25. Zuliani, G and Hobbs, H H. (1990) A high frequency of length polymorphisms in repeated sequences adjacent to Alu sequences. *Am. J. Hum. Genet.* **46,** 963–969
26. Economou, E. P., Bergen, A. W., Warren, A. C., and Antonarakis, S. E. (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc Natl. Acad. Sci. USA* **87,** 2951–2954.
27. Chakraborty, R., Lidsky, A. S., Daiger, S. P., Guttler, F., Sullivan, S., Dilella, A., and Woo, S. L. C. (1987) Polymorphic DNA haplotypes at the human phenylalanine hydroxylase locus and their relationships to phenylketonuria. *Hum. Genet.* **76,** 40–46.
28 Chakravarti, A., Buetow, K. H , Antonarakis, S. E., Waber, P. G., Boehm, C. D., and Kazazian, H. H. (1984) Nonuniform recombination within the human beta-globin gene cluster. *Am. J. Hum. Genet.* **36,** 1239–1258.
29 Estivill, X., Scambler, P. J , Wainwright, B J., Hawley, K., Frederick, P., Schwartz, M., Baiget, M., Kere, J., Williamson, R., and Farrall, M. (1987) Patterns of polymorphism and linkage disequilibrium for cystic fibrosis. *Genomics* **1,** 257–263.
30. Solomon, E. and Bodmer, W. F. (1979) Letter to the editor. *Lancet* **1,** 923.
31 Botstein, D., White, R., Skolnick, M , and Davis, R. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms *Am. J. Hum. Genet* **32,** 314–331.
32. Human Gene Mapping Workshop (1974) First international workshop on human gene mapping. *Cytogent. Cell Genet.* **13,** 1–216.
33. Human Gene Mapping Workshop (1989) Tenth international workshop on human gene mapping. *Cytogent. Cell Genet.* **51,** 1–1147.
34. Lathrop, G M., O'Connell, P., Leppert, M , Nakamura, Y., Farrall, M., Tsui, L.-C., Lalouel, J -M , and White, R. (1989) Twenty-five loci form a continuous linkage map of markers for human chromosome 7. *Genomics* **5,** 866–873.
35 Donis-Keller, H., Green, P., Helms, C , Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T. P , Bowden, D. W., Smith, D. R., Lander, E. S., Botstein, D., Akots, G , Rediker, K. S , Gravius, T., Muller-Kahle, H., Fulton, T. R., Ng, S , Schumm, J W., Braman, J. C., Knowlton, R G., Barker, D. F., Crooks, S. M., Lincoln, S., Daly, N., and Abrahamson, J. (1987) A genetic map of the human genome. *Cell* **51,** 3197.
36 White, R., Lalouel, J -M., O'Connell, P., Nakamura, Y, Leppert, M., and Lathrop, M. (1987) Linkage map of human chromosomes (Howard Hughes Medical Institute, Salt Lake City, UT).
37 Sherrington, R., Brynjolfsson, J., Pertursson, H., Potter, M., Dudleston, K., Barraclough, B , Wasmuth, J., Dobbs, M , and Gurling, H. (1988) Localisation of a susceptibility locus for schizophrenia on chromosome 5 *Nature* **336,** 164–170.
38 McCarthy, T V., Healy, S. J M., Heffron, J. J. A., Lehane, M., Deufel, T., Lehmann-Horn, F., Farrall, M., and Johnson, K J (1990) Localisation of the malignant hyperthermia susceptibility locus to human chromosome 19q12-q13.2. *Nature* **343,** 562–564.
39 Jeffreys, A J , Wilson, V , Thein, S. L., Weatherall, D. J., and Ponder, B. A. J. (1986) DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *Am. J Hum. Genet.* **39,** 11–24.
40 Chamberlain, S , Worrall, C. S., South, S., Shaw, J., Farrall, M., and Williamson, R. (1987) Exclusion of the Friedreich's ataxia gene from chromosome 19. *Hum. Genet.* **77,** 122–126
41 Wallis, J., Shaw, J., Wilkes, D , Farrall, M., Williamson, R., Chamberlain, S., Skare, J C , and Milunsky, A. (1989) Prenatal diagnosis of Friedreich's ataxia. *Am J Med Genet.* **34,** 458–461.

42. Ott, J. (1974) Estimation of the recombination fraction in human pedigrees· Efficient computation of the likelihood for human linkage studies. *Am. J. Hum. Genet.* **26**, 588–597.
43. Lathrop, G. M., Lalouel, J.-M., Julier, C., and Ott, J. (1984) Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* **81**, 3443–3446.
44. Lander, E. S., Green, P., Abrahamson, J., Barlow, A , Daly, M. J., Lincoln, S. E , and Newburg, L. (1987) Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181.
45. Morton, N. E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum Genet.* **7**, 277–318.
46. Klinger, K. W., Stanislovitis, P., Hoffman, N., Watkins, P. C., Schwartz, R., Doherty, R., Scambler, P., Farrall, M., Williamson, R , and Wainwright, B J. (1986) Genetic homogeneity of cystic fibrosis. *Nucleic Acids Res.* **14**, 8681.
47 Egland, J. A., Gerhard, D S., Pauls, D. L., Sussex, J N , Kidd, K. K., Allen, C R , Hostetter, A. M., and Housman, D. E. (1987) Biopolar affective disorders linked to DNA markers on chromosome 11. *Nature* **325**, 783–787.
48 Kelsoe, J R., Ginns, E. I , Egeland, J. A., Gerhard, D S , Goldstein, A. M , Bale, S. J., Pauls, D. L., Long, R. T., Kidd, K. K., Conte, G., Housman, D. E., and Paul, S M (1989) Re-evaluation of the linkage relationship between chromosome 11 p loci and the gene for bipolar affective disorder in the Old Order Amish. *Nature* **342**, 238–243
49. Robertson, M. (1989) False start on manic depression. *Nature* **342**, 222.
50. Ott, J. (1985) *Analysis of Human Genetic Linkage* Johns Hopkins University Press, Baltimore, MD.
51. Conneally, P. M., Edwards, J H., Kidd, K. K., Lalouel, J.-M , Molton, N. E., Ott, J., and White, R. (1985) Report of the committee on methods of linkage analysis and reporting. *Cytogenet. Cell Genet* **40**, 356–359.
52. Boehnke, M. (1986) Estimating the power of a proposed linkage study; a practical computer simulation approach. *Am J. Hum Genet.* **39**, 513–527
53. Morton, N. E. (1956) The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am. J Hum. Genet.* **8**, 80–96.
54. Risch, N. (1990) Linkage strategies for genetically complex traits. I, II and III. *Am J Hum Genet.* **46**, 222–253.
55. Bishop, D. T. and Williamson, J. A. (1990) The power of identity-by-state methods for linkage analysis. *Am J. Hum. Genet.* **46**, 254–265.

CHAPTER 30

Diagnosis of Genetic Disorders with Linked DNA Markers

Christopher G. Mathew

1. Introduction

The development of techniques for the analysis of specific DNA sequences has led to the discovery of a vast amount of variation of DNA sequence among different individuals. Consequently, it is now usually possible to distinguish the two parental copies of a particular chromosomal region in an individual. The difference arises either from the presence or absence of a restriction enzyme site in the region, or from a difference in the number of tandemly repeated sequences present in the two alleles. Such differences were originally detected as variations in the length of restriction fragments (restriction fragment length polymorphisms or RFLPs) after blotting and hybridization with probes for unique sequences in the region (*see* Chapter 15), but are now increasingly being detected by means of the polymerase chain reaction or PCR (*see* Chapters 1 and 6).

If the DNA sequence polymorphism occurs within or close to a gene that is mutated in an individual, it can be used to trace the inheritance of the mutant gene in his or her offspring. In Fig. 1, for example, an individual who is heterozygous for a DNA polymorphism with alleles A1 and A2 and who carries a mutation in a gene nearby, produces an affected child who has inherited the disease gene together with allele A1. The unaffected parent is homozygous for the A1 allele. Future offspring who inherit the A1 allele from their affected parent are also likely to be affected since the mutation and the A1 allele are unlikely to be separated from each other by recombination dur-

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C. Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

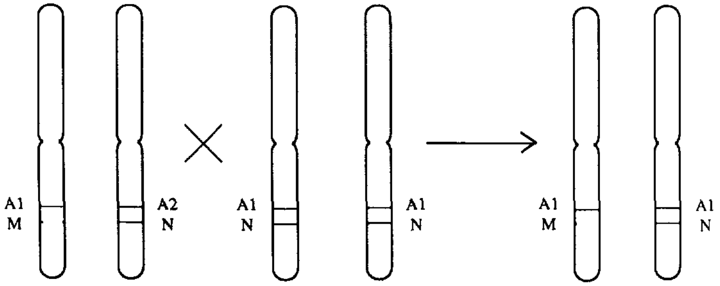


Fig. 1. Schematic of a pair of chromosomes from members of a family with a genetic disorder for which the gene is tightly linked to a DNA polymorphism with alleles A1 and A2. M = mutant copy of the gene, and N = normal copy.

ing meiosis. Thus, once the linkage phase has been established in the family, i.e., knowing which of the two alleles cosegregates with the mutated gene, the information can then be used for predicting whether other offspring or a fetus will be affected.

The advantage of this approach of diagnosis by "linkage" of an allele of a polymorphism to the mutation is that knowledge of the causative mutation in the family is not required. Furthermore, the diagnosis can be done even if the gene responsible for the disorder has not yet been isolated, provided that a polymorphism closely linked to the disease gene is available. In Huntington's disease, for example, the location of the gene responsible was mapped using the approach described in Chapter 29, and many prenatal diagnoses have been done, but the gene responsible for the disorder has not yet been isolated. The disadvantage is that a family study is required to establish the linkage phase for the DNA marker, and samples from key individuals may not be available. Also, the accuracy of the diagnosis is dependent on the frequency of recombination between the marker and the mutation.

The purpose of this chapter is to describe the strategy to follow for the diagnosis of inherited disorders with linked DNA markers, to outline applications of this approach, and to discuss some of the complications that may be encountered. Details of the molecular protocols used may be found in Chapters 6 and 15.

2. Strategy

2.1. Establish the Correct Clinical Diagnosis

This may seem self-evident, but the clinical phenotype of different genetic disorders may overlap. For example, a child with an autosomal recessively inherited limb girdle dystrophy may initially be diagnosed as an X-linked Duchenne muscular dystrophy (DMD). Neurofibromatosis type 1, which maps

on chromosome 17, has features in common with neurofibromatosis type 2 from chromosome 22. It is therefore important that the family be assessed by a physician with the requisite experience of the disorder.

2.2. Choose Suitable DNA Markers

2.2.1. Is the Marker Tightly Linked to the Disease Gene?

If the DNA polymorphism lies within the disease gene, it is likely to be tightly linked with a recombination frequency of less than 1%. However, there are exceptions. The dystrophin gene, which is mutated in DMD, is very large, and shows 12% recombination between its 5' and 3' ends *(1)*. If linkage is at a recombination frequency of 1–2% or less, a single marker on one side of the gene is generally adequate to provide a diagnosis; patients will usually accept a 1–2% risk of error from recombination between the marker and the mutation if a direct test for the mutation is not available. If the recombination frequency is closer to 5%, it is advisable to use two markers, one on either side of the gene, since it is very likely that a recombinant would be detected (*see* Section 3.3 for an example). If no recombination is apparent between the flanking markers, misdiagnosis would only result if each of the flankers had recombined with the mutation. The probability of this occurring is the product of the individual recombination frequencies, which is very low (e.g., $0.05 \times 0.05 = 0.0025$, or less than 1%).

2.2.2. Is the Marker Likely to Be Informative?

The affected parent in the family must be heterozygous for the polymorphism in order to establish which of the two copies of that chromosomal region his or her offspring have inherited. The majority of RFLPs result from single nucleotide substitutions, and therefore produce two possible alleles. Consequently, at most, 50% of the population could be heterozygous for the marker. The other class of markers, which result from variable numbers of tandem repeats at a locus (VNTRs or minisatellites, *see* refs. *2,3*), are more informative since multiple alleles exist in the population, and heterozygote frequencies of 70–80% are common. Thus, VNTR markers should be chosen wherever possible, or two-allele markers with the highest heterozygosity. If the polymorphisms are to be detected by blotting, and a range of RFLPs are available, choose those that are detected with the same enzyme where possible, since a single filter can be reprobed many times.

2.2.3. Can the Marker Be Amplified by PCR?

The DNA sequence around many RFLPs has now been established, and specific oligonucleotide primers developed to allow amplification of these sequences by PCR (*see* Chapter 6). This has produced great savings in time and cost, and radioactivity is not required for the analysis. Furthermore, a

new class of VNTR polymorphims have been discovered, which are simple tandem repeats, such as $(CA)_n$ or microsatellites *(4,5)*. These repeats have multiple alleles, high heterozygosities and are analyzed by PCR. They appear to be ubiquitous in the human genome, and will be used increasingly for DNA diagnosis in the future. Microsatellites are usually analyzed by using ^{32}P-labeled primers in the PCR; the alleles are resolved on polyacrylamide sequencing gels, and detected by autoradiography *(4,5)*. Recently, however, it has been shown that the alleles can be resolved on nondenaturing polyacrylamide gels and detected using a silver stain *(6)*.

2.3. Check the Accuracy of Your DNA Analysis

When the analysis of a polymorphism is being set up in a laboratory, whether by blotting or by PCR, it is important that staff ensure that they can produce reliable results before the test is put to diagnostic use. For example, a set of samples that has been typed for the marker by another laboratory can be obtained and tested "blind" for concordance. If this is not possible, the marker should at least be checked for correct Mendelian segregation in a set of families, and shown to be linked to the appropriate disease gene in individuals whose diagnosis has already been established. Several common pitfalls that may lead to mistyping by blotting or PCR are listed in the Notes (Section 6, 1–6).

2.4. Analysis of a Family

2.4.1. Who Wants to Know What?

The first step in linkage analysis of a family is to establish clearly, with the help of clinical colleagues who have counseled the family, what the analysis is expected to achieve. For example: does the son want to know his carrier risk, or would the eldest daughter, who is getting married, like prenatal diagnosis? Once this is established, it will be possible to decide whether one has the necessary samples to achieve these objectives, or whether further family members will have to be sampled.

2.4.2. Linkage Analysis

The DNA markers that are available can be given a priority rating based on whether they can be analyzed by PCR, the tightness of their linkage to the disease gene, and their heterozygosity. All family members should be typed with the best set of markers, as this is generally more efficient than typing only the affected parent and having to go back and type the rest of the family if the marker is informative. Suitable controls should be included on blots or in PCRs (*see* Notes 1 and 4), and the results should be checked by an experienced member of the laboratory staff.

The results can now be analyzed on the pedigree. If fully informative (*see* examples in Section 3), the report can be written. If not, the "second string"

markers are then typed. If the family is still uninformative, the options are to develop further markers if available, refer the family to a specialist laboratory that has a wider range of markers, or to report that the family is uninformative and put them on hold until new information becomes available.

2.5. Prenatal Diagnosis

2.5.1. Preparation

Prenatal diagnosis should normally only be undertaken if a linkage study has already been done on the family and if informative markers are available. "Crash" pregnancies, in which the mother is already pregnant but the family has not been analyzed, are very time consuming, since other work has to be suspended while the "crash" family is typed with all available markers. When the laboratory has been informed that a prenatal diagnosis has been scheduled for a particular family, staff should check that all reagents required for the analysis are available, and that all probes or PCRs to be used are working well.

2.5.2. DNA Extraction

The fetal sample provided is usually a chorionic villus biopsy. The sample should first be freed of any contaminating maternal tissue using a dissecting microscope. Alternatively, cells from an amnion or placental culture may be used. DNA can then be extracted using protocols described in a previous volume in this series *(7)*. Because of the great sensitivity of PCR, it is now possible to analyze small quantities of biopsy tissue or uncultured amniotic fluid. Boil 1–2 mg of tissue in 50 µL of TE buffer (10 m*M* Tris-HCl, 1 m*M* EDTA, pH 8) for 15 min in a microfuge tube, centrifuge for 1 min and use 5–10 µL of the supernatant for the PCR. Amnion cells can be pelleted from 5 mL of fluid and boiled as for the tissue. If these procedures are to be used, trial runs should be done before attempting a prenatal diagnosis (*see* Note 7).

DNA from the fetus and from other key family members is then typed with the relevant markers, including the appropriate controls (*see* Notes 1 and 4). Once a result is given, the pregnancy should be followed up to establish the outcome. For example, if a low risk prediction is given and the pregnancy continued, clinical tests may be carried out shortly after birth to confirm that the child is unaffected.

3. Applications

3.1. Autosomal Dominant Disorders

The aim of linkage analysis in these disorders is to establish whether individuals who are at risk for a disorder are likely to have inherited the mutation, and to offer prenatal diagnosis to those that have. For some disorders, such as the inherited cancer syndrome multiple endocrine neoplasia type 2A

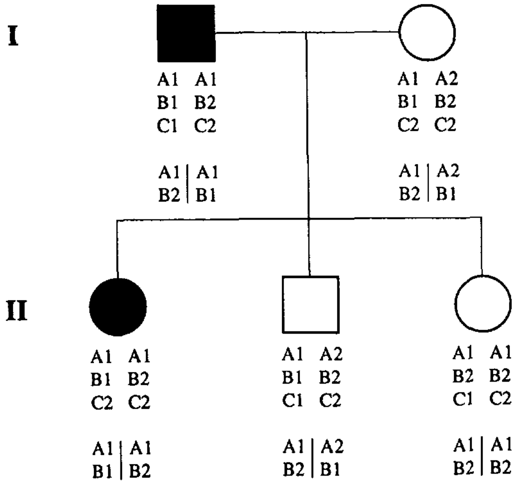


Fig. 2. Family with an autosomal dominant genetic disorder, who have been typed with three linked DNA markers A, B, and C. Haplotypes are indicated for the markers A and B below the genotypes for all thee markers. Filled-in circles and squares represent affected females and males, respectively. Empty symbols indicate clinically unaffected individuals who may or may not be gene carriers.

(7), presymptomatic testing is obviously clinically valuable, since family members who have inherited the high risk allele will be subjected to intensive screening for tumors followed by surgery as soon as they appear. For late onset disorders, such as Huntington's disease, at risk individuals will request linkage analysis both to establish whether they are likely to be affected and, if they are carriers, for prenatal diagnosis.

An example of linkage analysis in a family with an autosomal dominant disorder is shown in Fig. 2. If only the "A" marker had been typed in the family, it would be uninformative, since the affected parent is homozygous. The "B" marker is also uninformative in this family since although the affected parent is heterozygous, the unaffected parent is also heterozygous, and the linkage phase cannot be established. The "C" marker is fully informative; the mutation segregates with the paternal C2 allele. Note that although markers A and B are uninformative if used alone, information from the two markers can be combined to construct the parental haplotypes for them. Thus, since individual II-3 is A1A1, B2B2, she must have inherited an A1B2 chromosome from each parent. The other chromosomes for I-1 and I-2 can therefore be deduced as A1B1 and A2B1, respectively. The haplotypes constructed from the genotypes of the offspring (*see* Fig. 2) show that II-2 and II-3 are at low risk for the disorder.

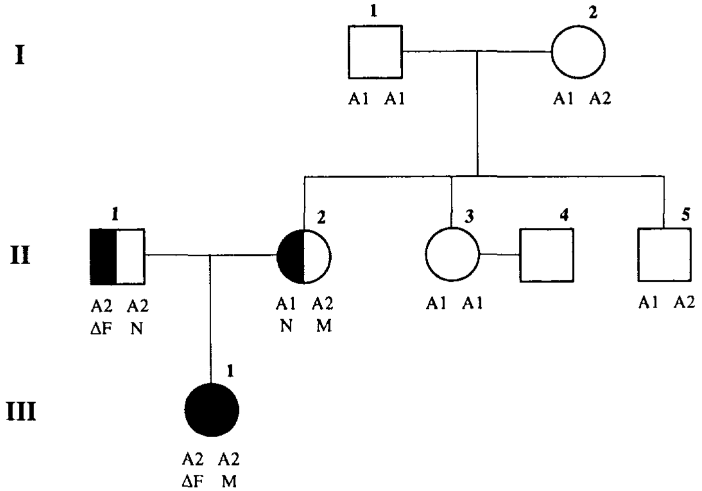


Fig. 3. Cystic fibrosis family typed with the linked marker A. ΔF represents the common CF mutation ΔF508. M = an undefined CF mutation, and N = a normal copy of the CF gene. Half filled-in circles and squares represents known CF carriers.

3.2. Autosomal Recessive Disorders

The principle is the same as for dominant disorders, but both parents must be heterozygous for a linked marker since both are carriers of a mutant gene, and each mutant must be traced in the offspring. An example of linkage analysis in a family with cystic fibrosis (CF) is shown in Fig. 3. The maternal CF gene is linked to an A2 allele, which was inherited from individual I-2. Thus, II-3, who would like to start a family, has a reduced carrier risk, but II-5 has a raised carrier risk. In this family, linkage analysis has been combined with mutation analysis. Although the father of the affected child is not informative for the linked marker, he has been found to carry the common CF mutation ΔF508 *(8)*. The maternal CF gene can be tested for in future pregnancies using the A marker, and the paternal gene tested for using a PCR assay for ΔF508 *(9)*.

3.3. X-Linked Recessive Disorders

The most commonly encountered disorders in this category are the Duchenne or Becker muscular dystrophies (DMD/BMD) and the hemophilias. The great majority of females who carry such mutations are asymptomatic. Males are hemizygous for the markers and do not inherit a paternal allele. This seems a very obvious statement, but it is surprising how often one slips into "autosomal mode" when analyzing these pedigrees. The objectives of linkage analysis are to use suitably linked X chromosome mark-

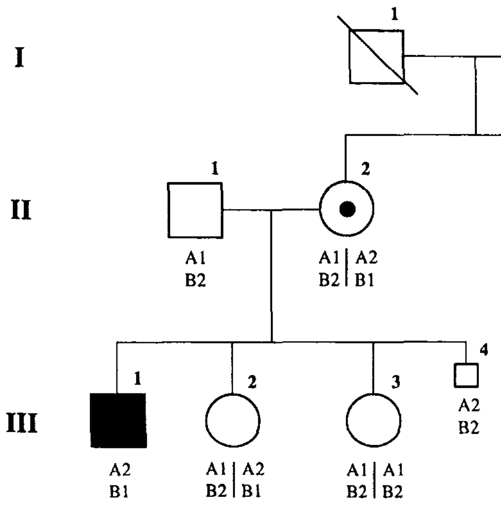


Fig. 4. Family with the X-linked disorder Duchenne muscular dystrophy, typed with markers at the 5' (A) and 3' (B) ends of the dystrophin gene. Diagonal lines indicate that an individual has died. The dot in the center of a female symbol indicates a known DMD carrier. The small square of individual III-4 indicates a male fetus.

ers to determine carrier risks for females and to perform prenatal diagnosis on male fetuses. This is generally relatively straightforward for the hemophilias, since either the mutations have been defined (*see* for example, Chapter 3) or tightly linked markers are available. Analysis of DMD families is much more complex in spite of the fact that the coding regions of the gene (dystrophin) have been cloned and about 60% of patients have gene deletions. This is partly because the gene has a very high mutation rate, so that about one third of cases arise from new mutations and most mutations are likely to be unique to a particular family. Furthermore, the gene is very large and has a total intragenic recombination frequency of about 12% *(1)*. Thus, a single intragenic marker does not provide a sufficiently accurate predictive test.

An example of linkage analysis in a DMD family is shown in Fig. 4. The "A" marker is located at the 5' end of the dystrophin gene, and the analysis has raised the carrier risk of individual III-2 since she has inherited the same maternal X at this locus as her affected brother (assuming paternity is correct—*see* Section 5.). However, since the location of the mutation within the gene is unknown in this family, the recombination rate between it and marker A could be up to 12%. Furthermore, there could be a crossover in either III-1 or III-2, so the degree of uncertainty in the diagnosis is unacceptably high.

The family was therefore typed with marker "B", which is located at the 3' end of the gene. Now it can be seen that III-2 inherited the high risk alleles for markers at both ends of the gene (maternal haplotype A2B1). She could only have failed to inherit the DMD mutation if each of the markers had recombined with the mutation, which has a probability of less than 1%. Her sister, (III-3), has the low risk A1B2 haplotype, and is therefore very unlikely to be a carrier. Finally, the male fetus in this family (III-4) has the haplotype A2B2, which suggests that an intragenic recombination event has occurred. Since we do not know whether the DMD mutation in this family is closer to the 5' or 3' end of the gene, we cannot offer a prediction for this fetus.

Note that it is particularly important to "cover" the dystrophin gene with markers when one is offering a low risk prediction, since such individuals will probably assume that they are free of the disease and that they do not require prenatal diagnosis. This argument applies within even greater force if one is using the markers for prenatal diagnosis.

A bonus for linkage analysis in DMD families is that if a deletion has been detected in the affected male and a female relative is heterozygous for a polymorphism within the deleted region, then she cannot be a carrier of that deletion (unless she is a germinal mosaic, *see* Section 5.2).

4. Risk Calculations

In straightforward cases, carrier risks or the risk of having an affected fetus can be calculated simply as the probability of recombination between the informative linked marker and the mutation in the disease gene. In Fig. 2, for example, let us suppose that the "C" marker recombines with a frequency of 2% with the mutation, and that we wish to calculate the carrier risk for individual II-2. The paternal C2 allele appears to be linked to the mutation. However, there is a 2% probability that a crossover occurred in II-1, in which case the high risk allele would be C1. A crossover could also have occurred in II-2. Thus, the probability of an error in the prediction for II-2 as a result of recombination is 2% + 2% = 4%, and his risk of carrying the mutation is approx 96%. In Fig. 4, the probability of error in prediction of carrier status for III-2 or III-3 is the product of the recombination rate between each of the markers and the mutation, which is $0.06 \times 0.06 = 0.0036$ or less than 0.4%. The carrier risk for III-2 is therefore greater than 99%, and for III-3, less than 1%.

Risk calculations based on DNA analysis can be combined with risks calculated on the basis of, for example, the age of the patient, or a biochemical assay, such as creatine kinase, which is elevated in two-thirds of DMD carriers. In such cases, Bayesian calculations can be used to calculate a combined risk that the patient carries the mutant allele (*see* Emery, ref. *10*, for examples).

If the structure of the kindred being analyzed is complex, it may be necessary to resort to computer programs to assist risk calculations. Programs, such as MLINK, which is part of the LINKAGE package, can be used for risk calculations. Further information on software is given in Chapter 31. The danger of using such programs is that a single error in data entry can result in wildly improbable risk results. Use of such programs should complement, but not replace, analysis of the pedigree using common sense.

5. Complications

5.1. Nonpaternity

Analysis of a kindred with a particular DNA marker may show, for example, that a child has a genotype A2A2, whereas the father is A1A1. In the absence of errors resulting from a mix-up in the labeling of tubes either in the clinic or in the laboratory, or from partial restriction enzyme digestions, this suggests that the stated father is not the biological father of the child. Undetected nonpaternity is quite likely to lead to errors in diagnosis by linkage analysis. For example, if the real father of individual III-2 in Fig. 4 was A2B1, then her maternal X chromosome would be A1B2, and her carrier risk for DMD would be very low rather than very high.

Nonpaternity may not be apparent from the markers that have been typed in the family, since the majority of polymorphisms in diagnostic use are two-allele systems. If nonpaternity is suspected because of remarks made by a family member during counseling or because of unlikely crossover events, or if the diagnosis is dependent on correct paternity, this should be checked. This can be done quite easily by hybridizing one of the highly polymorphic minisatellite probes developed by Alec Jeffreys *(2)*, to blots of the family's DNA cut with a suitable restriction enzyme *(11)*. Since these probes detect multiple alleles, each of which is present at a low frequency in the population, use of a single locus-specific minisatellite probe is usually sufficient to detect nonpaternity.

5.2. Germinal Mosaicism

Families have been reported with disorders, such as DMD, in which a mutation has been transmitted to more than one offspring by a parent who does not have the mutation in their own somatic cells *(12)*. This indicates that the mutation occurred early in the proliferation of the germline, leading to germinal mosaicism. This phenomenon has important implications for linkage diagnosis; for example, a woman who is heterozygous for an RFLP located within the region deleted in her affected son would be diagnosed as a noncarrier (*see* Section 3.3). However, if the deletion is present in a significant proportion of her germ cells, she may have another affected son. The

existence of this phenomenon may persuade women with a single affected child and no family history to opt for prenatal diagnosis even if they do not carry the mutation in their somatic cells.

5.3. Genetic Heterogeneity

In some genetic disorders, the mutant genes responsible may be located on more than one chromosome. Genes for tuberous sclerosis have been localized by linkage analysis to chromosomes 9 and 11, although the clinical symptoms of the two groups are similar *(13)*. If linkage analysis is being done in a family with such a disorder, markers from both loci will have to be typed in order to establish to which group that particular family belongs. If the family is small, this may be difficult or even impossible to determine.

6. Notes

1. Partial digestion of genomic DNA or PCR products with restriction enzymes may lead to incorrect typing, such that an individual who is actually homozygous for the smaller allele (i.e., A2A2) appears to be A1A2, or an A1A2 genotype appears to be A1A1. Partial digests of genomic DNA can usually be detected on the stained gel since the average size of the restriction fragments will be larger than those in the fully digested lanes. Also, some restriction enzymes, such as EcoRI and BamHI, cut within highly repeated satellite DNA sequences, and produce discrete bands of a characteristic size that can be seen in the gel. In digests of PCR products, particular attention should be paid to the intensity of the stained DNA fragments. The larger fragments should stain brighter than the smaller ones since they contain more DNA. It is advisable to include DNA from a known heterozygote in every blotting or PCR experiment.
2. If the hybridization signal from probing a blot is weak, or if a low yield has been obtained from a PCR, the less intense of two alleles may not be seen. If the signal from the "stronger" of the two alleles is faint, the test should be repeated.
3. Contamination of a human DNA sample or digest with a plasmid DNA can result in the appearance of spurious bands on autoradiographs of the blots. This is because most probes will be contaminated with traces of the plasmid vector in which they were cloned, and most commonly used plasmids are related. The spurious band may be of the same size as an allele of an RFLP, and thus lead to an incorrect diagnosis. Such contaminants can often be spotted because they produce bands of greater intensity than expected. They can be detected by hybridization of the filter with a probe consisting of the relevant plasmid vector only.

4. Contamination of a genomic DNA sample or one of the solutions used for PCR with product from a previous PCR is an important potential source of error. Since this is a pure DNA sequence, it need be present in only very small quantities for it to be amplified to the same extent as the corresponding sequence in the sample being analyzed. Contamination can be minimized by use of a separate set of automatic micropipets and tips for PCR products, and a separate set of solutions for setting up the PCR. A negative control that contains all the reaction components except sample DNA should be included in every set of PCRs.
5. Nonspecific amplification of DNA from other regions of the genome could result in spurious bands that comigrate with genuine alleles of an RFLP. Reaction conditions, such as the annealing temperature, can be manipulated until a "clean" PCR is obtained (*see* Chapter 1).
6. NAFNAP is a term coined by Abbs et al. *(14)* to describe nonamplification resulting from nonannealing of a primer. This can occur if a sequence polymorphism is located within the primer binding site, which destabilizes primer binding at the annealing temperature sufficiently to prevent amplification. Thus, the allele at which the undetected polymorphism occurs will not be amplified, and the individual will be mistyped. The defense against this possibility is to compare a substantial number of PCR results for an RFLP with those obtained on blots. If this cannot be done, the RFLP can be typed with an alternative pair of primers, and the results compared with those obtained in the original PCR.
7. Crude DNA preparations should be stored frozen as they are more susceptible to degradation by endonucleases than those produced by extraction with organic solvents.

References

1. Abbs, S., Roberts, R. G., Mathew, C G , Bentley, D R., and Bobrow, M. (1990) Accurate assessment of intragenic recombination frequency within the Duchenne Muscular Dystrophy gene. *Genomics* **7**, 602–606
2. Wong, Z., Wilson, W., Patel, I., Povey, S., and Jeffreys, A. J. (1987) Characterisation of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.* **51**, 269–288.
3. Nakamura, Y, Leppert, M., O'Connell, P , Wolff, R , Holm, T., Culver, M , Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**, 1616–1622.
4. Weber, J. L. and May, P. E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396.

5. Litt, M. and Luty, J A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44,** 397–401.
6. Love, J. M., Knight, A M , McAleer, M. A., and Todd, J. A. (1990) Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucleic Acids Res.* **18,** 4123–4130.
7. Mathew, C. G. P., Easton, D. F., Nakamura, Y., Ponder, B. A. J., and members of the MEN2A study group (1991) Presymptomatic diagnosis of multiple endocrine neoplasia type 2A using linked DNA markers *Lancet* **337,** 7–11.
8 Riordan, J. R., Rommens, J. M , Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielinski, J., Lok, S., Plasvic, N , Chou, J.-L., Drumm, M. L., Ianuzzi, M. C., Collins, F S., and Tsui, L.-C. (1989) Identification and cloning of the cystic fibrosis gene· Cloning and characterisation of the complementary DNA. *Science* **245,** 1066–1073.
9. Mathew, C. G. P., Roberts, R , Harris, A., Bentley, D. R., and Bobrow, M. (1989) Rapid screening for the ΔF508 deletion in cystic fibrosis. *The Lancet* **ii,** 1346.
10. Emery, A. E. H. (1990) Bayesian methods in medical genetics, in *Principles and Practice of Medical Genetics,* vol. 1 (Emery, A. E. H. and Rimoin, D. L., eds.), Churchill Livingstone, Edinburgh, London, Melbourne, and New York, pp. 107–113.
11. Telenius, H., Clark, J., Marcus, E., Royle, N., Jeffreys, A J., Ponder, B. A. J., and Mathew, C G P (1990) Minisatellite DNA profiles: Rapid sample identification in linkage analysis. *Hum. Hered.* **40,** 77–80.
12. Bakker, E., Van Broekhoven, Ch., Bonten, E. J. van de Vooren, M. J., Veenema, H., Van Hul, W., Van Ommen, G. J. B., Vandenberghe, A , and Pearson, P. L. (1987) Germinal mosaicism and Duchenne muscular dystrophy mutations. *Nature* **329,** 554–556.
13 Janssen, L. A. J., Sandkuyl, L. A., Merkens, E. C., Maat-Kievit, J. A., Sampson, J. R., Fleury, P., Hennekam, C M , Grosveld, G. C., Lindhout, D., and Halley, D. J. J. (1990) Genetic heterogeneity in tuberous sclerosis. *Genomics* **8,** 237–242.
14. Abbs, S., Yau, M., Clark, S , Mathew, C. G. P., and Bobrow, M (1991) A convenient multiplex PCR system for the detection of Dystrophin gene deletions: A comparative analysis with cDNA hybridisation reveals mistypings by both methods *J. Med. Genet* **28,** 304–311

Chapter 31

Software for Genetic Linkage Analysis

Stephen P. Bryant

1. Introduction

Gene mapping by the analysis of traits segregating in human pedigrees is a major goal of linkage analysis *(1)*, itself firmly rooted in the statistical technique of maximum-likelihood estimation (MLE; *2*). The quantity estimated is most often the recombination fraction (θ), using the now well-known lod-score method (ref. *3* and Chapter 29, this volume).

The use of MLE has been facilitated by the development of algorithms *(4)* that can be implemented on small computers in such packages as Liped *(5)* and Linkage *(6)*. New, more efficient algorithms to perform multipoint linkage analysis have recently appeared *(7)* and have extended the size of map that can realistically be created using the method *(8)*.

With these algorithms, finding genetic linkage to a putative disease susceptibility locus demands the use of a suitable transmission model, which may require the joint estimation of several parameters. With traditional MLE procedures, this can be difficult. The affected sib-pair method *(9)*, of immense value in the analysis of the human leukocyte antigen (HLA) region, is based on the concept of identity by descent. Lange *(10)* applied the affected-sib method to sib-sets, and later to identity by state *(11)*. Weeks and Lange *(12)* generalized the method to extended pedigrees. They used the algorithm of Karigl *(13)* to compute multiple-person kinship coefficients and thence to derive the distribution of a test statistic within each pedigree. Their Kin package *(12)* enables tests of the hypothesis of Mendelian segregation among related, affected individuals.

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by: C Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

Weitkamp and Lewis *(14)* used Monte Carlo simulation methods in conjunction with an identity-by-descent statistic to test for Mendelian segregation in extended pedigrees. Their Pedscore program is similar to Kin in spirit, though it cannot be applied to identity by state.

The original use of simulation on genealogies was described by Edwards *(15)*. MacCluer et al. *(16)* simulated gene flow through a genealogy, and computer simulation is also used in programs like Simlink *(17)* to estimate the power of a proposed linkage study. Ott *(18)* considered simulation methods applied to problems of linkage and heterogeneity, which will surely result in valuable software developments.

The development of software for genetic analysis encompasses several related issues. One consideration is that the design should enable integration with existing databases and applications. "Metadictionaries" (*see* Note 4) and thesauri *(19)* offer a promising approach to the integration of knowledge from diverse sources. Protocols have been developed *(20)* for the exchange of data between disparate genetic and molecular biological databases. Such paradigms as entity–relationship–attribute analysis *(21)*, relational synthesis *(22,23)*, and implementation under a relational database management system (RDBMS) facilitate software integration and the exchange of data between databases. It has to be said that these techniques are still not universally adopted, and exchange of data between applications is distinctly nontrivial. I have indicated where recognized paths exist between applications (Table 1).

The discipline of knowledge engineering and the appearance of intelligent knowledge-base management systems (IKBMS) will support and enhance the process. Expert systems (ES) have already shown promise in other areas of human genetics *(24)*. The utility of analytical software is extended by ancillary packages that are geared toward data management *(25)* and such support roles as pedigree display *(26,27)*.

Most software described here is freely available from either the originators of the package or, in some cases, from third parties, though it is usually not in the public domain. Users are advised that this is not always the case, and to register with distributors for the timely receipt of upgrades and bug fixes.

2. Hardware

The IBM PC family, principally the AT and compatibles, has been the mainstay of the linkage analyst since the early 1980s. Since 640 kilobytes (kB) are sufficient to handle mapping problems of moderate size, these machines are undoubtedly very cost-effective. Machines with much less than 512 kB will probably lack sufficient power to run anything but the smallest two-point analyses. A monochrome display is fine for most programs. As an exception,

Table 1
Software for Linkage Analysis—Technical Notes

Package	References	Version[a]	Language[b]	OS[c]	Export[d]
Linkage Program Package LPP	*6, 29–31*	4.9	Pascal, C	DUV	–
Liped	*5, 32*	November 1987	Fortran	DUV	–
Cri-Map	–	2.2	C	UV	–
Mapmaker	*27*	1.0	C	UV	–
MAP	*40*	–	Fortran	–	–
Mendel, Dgene, and Fisher	*39*	2.3, 1.0, 2.1	Fortran 77, Dbase III+	D	–
Linksys	*28*	4.11	Pascal	D	Linkage, Liped.
Simlink	*17, 34*	4.0	Fortran	D	–
Patch	*35*	11	C	D	–
Shell	–	1.0	Oracle, Pascal	V	Linkage, Liped, Kin, Pedigree/Draw.
Kin	*12*	1.0	Pascal, Fortran	DU	–
Pedigree/ Draw	*25*	4.0	Fortran	M	–
Pedpack	*36, 37*	3.0	C	U	Cri-Map
Plot2000	–	3.0	DBase III+	D	–

[a]The latest known version on general release. Prerelease (Beta) versions may be available for some packages

[b]This is the source code with which the package was developed It may or may not be available

[c]Operating System availability: D, DOS; U, UNIX; V, VMS; M, MAC

[d]An indication as to the facility to transfer data between packages.

Plot2000 *(27)* requires at least a color graphics adaptor (CGA) if genealogies are to be displayed on the screen.

The increasing availability of restriction fragment length polymorphism (RFLP) data over the last five years has resulted in the development of systems able to construct maps of upwards of 100 loci *(28)*. The algorithms that these systems use really need fast, 32-bit architecture if they are to be effec-

tive. The Sun or Apollo work stations, including the recent RISC-based Sparcstations and DEC stations, provide particularly good value. The DEC VAX minicomputer family running VMS or ULTRIX is a more traditional alternative for large problems.

The Apple Macintosh, though not a particularly powerful machine in terms of speed, has been gaining in popularity over the last few years, largely because of its ergonomic user interface. Packages such as Linkage and Liped should export easily with the right compiler, although versions are not generally available. It is likely that later versions of such packages as Mapmaker will be brought out for the Mac as well, once the necessity for 32-bit architecture is overcome. At the moment, the Macintosh is used in a support capacity *(26)*.

3. Operating Systems

The IBM PC and compatibles are almost always used with the DOS operating system; Version 2.1 and above is usually enough. OS/2 has still to gain much more than a token foothold in any PC area. VMS Version 5.1 for the DEC VAX is the latest version of that operating system and should be present on all VAX/VMS sites. UNIX is provided in the forms of SunOS or Apollo Domain for those particular machines and ULTRIX for the DEC VAX. The Macintosh has a full-fledged version of UNIX (A/UX).

4. Compilers

Most packages are distributed as source code with executable programs for one or, in some cases, several architectures. Some software is designed to be highly configurable and will need to be edited and recompiled at host sites. If changes to the source code are to be attempted, or if the code is to be exported to another, unsupported system, a compiler will be needed. Most software described in this report has been written in Pascal, C, or Fortran. Some operating systems are delivered with the compilers as standard features. This is true for most UNIX systems. In SunOS, the C compiler is supplied with the operating system, but the Pascal and Fortran compilers are available only at extra cost. Most VMS sites have C, Pascal, and Fortran, although these are not provided as standard features. DOS machines are supplied without compilers of any sort. Microsoft Pascal, Microsoft C, Microsoft Fortran, Turbo Pascal, and Turbo C will probably be the most useful.

5. DOS Emulators

Users should be aware of the existence of DOS emulators that enable the Macintosh or Sun user to execute DOS programs on their machines. These are usually slow, may require extra hardware, and cannot yet be rec-

ommended for serious use, given the availability of source code and the ease of portability.

6. Communications Software

A file-transfer and terminal-emulation package is essential if large problems are to be tackled. Kermit is still the most widely used of these packages. It is in the public domain and is distributed in the United Kingdom by the University of Lancaster for a variety of operating systems and architectures. Linksys *(25)* offers good integration with Kermit and is ideal for implementing a loose client–server strategy in which the PC is used as a tool for data management and preparation, and the actual numerical work is delegated to the powerful remote machine. Most of the software described below uses ASCII text files to store data. These can be transferred by Kermit or electronic mail, and will not vary from machine to machine.

7. Editors and Data-Management Aids

The only essential quality of an editor for linkage data is that it should be able to produce clean text files. A good, basic editor is included as part of the Linksys package *(25)*. Wordstar, Wordperfect, and Microsoft Word are proprietary systems that can be used to generate suitable text files.

8. Linkage Analysis and Ancillary Software

This section is, in a sense, a comparative review, since restrictions and positive features are evaluated and compared across systems. However, there are few directly competing packages: The programs described tend to complement each other. Hopefully this section will serve as a guide to the benefits of investing time to acquire and get to know a particular package. I have tabulated much of the basic information needed (Tables 1,2) and supplemented this with a short critical paragraph for each system (*see* Note 2).

8.1. Linkage Program Package (LPP)

LPP is a general-purpose package for multipoint linkage analysis, including risk calculation *(6,29–31)*. It includes modules for managing the data, preparing it for analysis, conducting the analysis, and interpreting the results.

Markers are divided into four types, which are sufficient to handle most kinds of genetic system likely to be encountered. These include codominant RFLPS, dominant–recessive systems, diseases with full or partial penetrance and general quantitative phenotypes. Data is prepared as text files, most easily with a data management aid like Linksys *(25)* or Shell. Versions of the

Table 2
Software for Genetic Linkage Analysis—Availability and Cost

Package	Type	Distributor	Cost[a]
Linkage Program Package LPP	Multipoint linkage analysis with risk calculation	Mark Lathrop CEPH 27 Rue Juliette Dodu 75010 Paris France	None
Liped	Two-point linkage analysis	Jurg Ott Columbia University Box 58 722 West 168 Street New York, NY 10032	None
Cri-Map	Multipoint map construction with some facility for disease loci	Phil Green Washington University Box 8232 4566 Scott Avenue St. Louis, MO 63110	None
Mapmaker	Codominant multipoint map construction in nuclear families	Mapmaker Distribution The Lander Lab Whitehead Institute for Biomedical Research Nine Cambridge Center Cambridge, MA 02142	Approx $50; variable
MAP	Multipoint map construction from two-point lod scores	Newton Morton Department of Community Medicine Southampton General Hospital University of Southampton Southampton England	None

(continued)

Table 2 *(continued)*
Software for Genetic Linkage Analysis—Availability and Cost

Package	Type	Distributor	Cost[a]
Mendel, Dgene, and Fisher	General genetic analysis, including data management	Daniel Weeks Department of Biomathematics UCLA School of Medicine Los Angeles, CA 90024-1766	None
Linksys	Data management for linkage and Liped	John Attwood Department of Genetics and Biometry University College London Wolfson House 4 Stephenson Way London NW1 2HE England	None
Simlink	Estimating the power of a proposed linkage study	Michael Boehnke Department of Biostatistics School of Public Health University of Michigan 109 South Observatory Ann Arbor, Michigan 48109	None
Patch	Haplotype deduction	Ellen M. Wijsman Medical Genetics, SK-50 University of Washington Seattle, WA 98195	None

(continued)

Table 2 *(continued)*
Software for Genetic Linkage Analysis—Availability and Cost

Package	Type	Distributor	Cost[a]
Shell	Data management for Liped, Linkage, and Kin	Stephen P. Bryant Human Genetic Resources Unit Imperial Cancer Research Fund Blanche Lane South Mimms Potters Bar Herts EN6 3LD England	None
Kin	Affected-pedigree-member method of linkage analysis	Daniel Weeks Department of Biomathematics UCLA School of Medicine Los Angeles, CA 90024-1766	None
Pedigree/Draw	Pedigree drawing	Jean W. MacCluer Department of Genetics Southwest Foundation for Biomedical Research PO Box 28147 San Antonio, TX 78284	None
Pedpack	General pedigree analysis and display	Alun Thomas School of Mathematical Sciences University of Bath Claverton Down Bath BA2 7AY England	£1000 UK

(continued)

Table 2 *(continued)*
Software for Genetic Linkage Analysis—Availability and Cost

Package	Type	Distributor	Cost[a]
Plot2000	Pedigree drawing	Don Bradley Institute of Medical Genetics University of Wales College of Medicine Heath Park Cardiff CF4 4XN Wales	£100 UK

[a]This should only be taken as a rough guide Contact the distributor to confirm costs

programs exist for linkage analysis of general (extended) pedigrees and of the nuclear families of the Centre d'Etude du Polymorphisme Humain (CEPH) collaboration.

LPP is distributed as Pascal (the core analytical programs) and C (the shell programs LCP and LRP) source code. The source code is highly portable and is available in UNIX, VMS, DOS, and generic formats. Source for LCP and LRP is not normally distributed, but is available on request. Executable programs are provided for DOS, VMS, or UNIX on a range of media, as required. The programs perform better with small numbers of markers (up to five for the general programs, substantially more for CEPH-style). The LPP package contains four programs: MLINK is used to construct two-point lod-score tables, LODCORE for iterative estimation of θ, ILINK for multipoint maps, and LINKMAP to insert markers into larger multipoint maps. It may be more practical to use Mapmaker (*see* Section 8.3) or Cri-Map (*see* Section 8.4) for large problems.

Separating the sexes is supported and interference can be accommodated. Data files are identical across all operating systems and all current versions.

8.2. Liped

Liped is a program for computing two-point lod scores in general pedigrees *(5,32)*. It handles markers using a phenotype–genotype matrix that is sufficiently general to code RFLPs and dominant–recessive systems. It handles age-dependent penetrance and division into liability classes. It also has the ability to separate the sexes.

Liped is distributed as Fortran source code. The code is reliable and easy to export to UNIX, VMS, or DOS. It can be obtained freely from the distributor or third parties, with restrictions (contact the distributor). Data files are ASCII text and can be easily generated using Linksys *(25)*.

8.3. Mapmaker

Mapmaker is an interactive package for the construction of codominant multipoint maps from nuclear CEPH-style families and F2 crosses *(28)*. It is not a general-purpose linkage package. It cannot be used for disease mapping and cannot be applied to extended pedigrees.

It is distributed in C source-code form. Executable versions are distributed for the VMS and UNIX operating systems. It cannot easily be exported to DOS or to the Macintosh. Utilities are provided to ease installation. UNIX versions have a "makefile" utility and VMS versions have a DCL script. For the next release of Mapmaker (Version 2), a new distribution and licensing procedure has been adopted. Contact the distributor for details.

Mapmaker uses an efficient algorithm for the computation of likelihoods, based on work done by Lander and Green *(7)*. Their expectation-maximization procedure requires fewer iterations, so a smaller number of computations are necessary to find the set of θs giving the map with the maximum likelihood. Even so, Mapmaker uses a vast amount of central processing unit (CPU) time. To put together a chromosome map of 45 markers over 50–60 families, as typically found within the CEPH collaboration, several hundred hours of CPU time can be involved. Under VMS, this may be deemed unacceptable by the local administration. It is possible to construct files for use in batch jobs, although all the interactive facilities of Mapmaker are then lost. It is advisable to submit these batch procedures to a queue with a high CPU limit (*see* Note 3).

A utility for exporting CEPH data to Mapmaker is available (*see* Note 1).

8.4. Cri-Map

Cri-Map is designed to facilitate the construction of large multilocus linkage maps. It was originally conceived to handle large numbers of codominant (RFLP) loci in CEPH-style nuclear families. Extensions to Version 2.2 enable it to be applied to certain types of disease loci and to handle general, extended pedigrees. It can cope specifically with those disease loci at which affected carriers are disallowed, that is, when full penetrance is assumed.

It is distributed as C source code. To use it, sites will need a compiler. The code follows closely the standards of Kernighan and Ritchie *(33)* and has been implemented under MicroVMS and ULTRIX by the original authors. I have exported it to SunOS UNIX. A "makefile" utility is available on request.

Its authors recommend that Cri-Map be used on machines with a minimum of 3 MB of available memory. In practice, we have found that for full chromosome data sets of 40–50 loci, 6–9 MB are necessary. It may be possible to run Cri-Map on an IBM PC for very small problems. Despite the large memory requirements, Cri-Map uses a very efficient algorithm for computing likelihoods. It was used in conjunction with Mapmaker to produce the Collaborative map of the complete human genome *(8)*. The price for efficiency is that Cri-Map uses less information from partially informative meioses than either Linkage or Liped. Population allele frequencies are not used in determining the relative probabilities of untyped founder genotypes. Some loci in untyped individuals are marked as uninformative and not subject to a full treatment. However, the information loss appears to be small. It uses a strategy different from that of Mapmaker in finding the maximum likelihood order.

A utility for exporting CEPH data to Cri-Map is available (*see* Note 1).

8.5. Kin

The theoretical background to Kin is given in Weeks and Lange (12). Building on the sib-pair method, they derived a statistic (a *Z*-score) that measures the similarity between typed, affected members of a pedigree on the basis of their marker genotypes. The distribution of *Z*-scores can be derived analytically or by simulation. *Z*-Scores are combined by Kin to give an overall *T* statistic that approximates a standard normal distribution. The extent to which the distribution of *T* approximates normality can be ascertained using Simulf, part of the Kin package.

Kin is supplied as source code in an IBM and UNIX version of Pascal. It is easy to export to VMS. Microsoft Pascal can compile the source code and is useful to optimize code against libraries and hardware. Simulf is written in Fortran. Executable code is provided for IBM PC compatibles with numeric coprocessors.

Certain types of family cause the program to crash, and these will have to be detected and removed by trial and error. The *T* statistic can be directly compared against a standard normal distribution using a one-tailed test. Kin does not use as much of the information available as would be used by Linkage and does not enable estimation of the recombination fraction. It is, however, extremely useful as a screening tool for linkage, since it does not require any assumptions to be made about the transmission model.

8.6. Simlink

Simlink is designed to estimate the power of a proposed linkage study based on a set of pedigrees of known structure and a genetic trait of interest *(17,34)*. It can handle both qualitative and quantitative traits, as well as sex and age-dependent penetrance.

It is distributed as a mixture of source and object code as well as an executable DOS program. It is actually based on a greatly extended version of Mendel (*see below*). It requires 640 kB of RAM and a numerical coprocessor. It is available without charge from the distributor.

8.7. Linksys

Linksys is the only generally available package for the management of genetic data to be used in conjunction with the analytical packages Linkage and Liped (*25*). It was written making heavy use of proprietary source code toolboxes, which are DOS-specific. It is unlikely that the present version could be exported to any other operating system.

Linksys manages data on pedigrees, markers, and phenotypes, and includes export facilities to Liped and Linkage. It performs as a shell to these systems, as well as to Kermit, enabling the user to prepare and execute analyses without being faced with a DOS command line.

Linksys is written in Turbo Pascal and distributed as executable code. Also supplied is a full-screen version of LCP (*see* Linkage Program Package LPP) that uses actual locus names instead of symbols.

Linksys is being substantially rewritten at the present time. The new Version 5 will incorporate a programming language based on Pascal, with which the user will be able to construct file-translation facilities from a library of user-supplied functions. This is a measure designed to ease the problems of data exchange between applications and databases. Also, data will be stored as indexed ASCII text files. This will make data recovery from corrupted media much more practical.

8.8. Pedigree/Draw

Pedigree/Draw is a general-purpose pedigree-drawing package for the Apple Macintosh (*26*). It can draw genealogies of almost any size, including those with some degree of inbreeding. It is used in conjunction with ARBOR, a program specifically designed to process complex, highly in-bred pedigrees into a form that can be displayed by Pedigree/Draw. Genealogies can be sent to a Postscript printer or imported into MacDraft or MacDraw for incorporation into figures.

It is distributed in executable form with plenty of example pedigrees and a good user guide. It is currently not available for any other operating system. It is obtainable without charge by writing to the distributor.

8.9. Patch

Patch is a program for deducing haplotypes from genotype data (*35*). It is distributed as C source code with DOS-executable programs. It is divided into modules for data collection and management, printing, and haplotype

deduction. The code is very portable and can probably be compiled onto most other architectures.

Patch can be obtained from the distributor, or from a third party on condition that the user register with Dr. Wijsman for bug fixes and other upgrades.

8.10. Pedpack

Pedpack Version 3.0 is a complete UNIX environment for pedigree analysis *(36,37)*. It offers facilities for checking, setting up, and editing pedigrees and genetic traits; probability and likelihood calculations; preparing data for the Pedigree Analysis Package (PAP—not considered here, but see ref. *38*), Linkage, and Cri-Map; computing gene-extinction probabilities by peeling and simulation; and drawing marriage node graphs.

The user requires some familiarity with the UNIX file structure and command language. It is distributed in a form suitable for Sun workstations running SunOS.

8.11. Mendel, Dgene, and Fisher

Mendel is a general modeling tool and can be used for segregation analysis, linkage analysis, and risk calculation without further modification *(39)*. It is supplied as a mixture of source and object code for the IBM PC and compatibles. Dgene manages data for export to Mendel and Fisher. It is supplied as DOS executable code, originally written in DbaseIII+. Fisher is designed to aid the epidemiological investigation of quantitative traits. It is supplied as a mixture of source and object code for the IBM PC and compatibles.

8.12. MAP

Map uses two-point lod scores to construct a multilocus map *(40)*. It can handle interference, but the main advantage is that lod scores from different sources can be combined irrespective of whether the raw data is available. It therefore uses a system entirely different from Cri-Map, Mapmaker, or Linkage. The two-point scores from these programs can be amalgamated and used as input to Map.

8.13. Shell

Shell is a system for managing data for the Linkage, Pedigree/Draw, and Kin packages. It was built using the tools created with the Oracle Database Management System. It has been run successfully on a VAX 8700 cluster running VMS 5.0 and Oracle 5.1.22. Shell is highly portable but sites will need an Oracle license. PC users will need expanded memory. It cannot, at present, be exported to the Apple Macintosh.

8.14. Plot2000

This is a program for managing and displaying genealogical data. Functionally it is very similar to Pedigree/Draw with the advantage that it runs on DOS machines with a minimum of 640 kB of RAM. To see the pedigrees, a graphics display (Hercules, CGA, EGA, or VGA) will be needed. Output to Epson-Style dot matrix printers is supported, as well as to those devices using HPGL. It is distributed as executable code for DOS. Earlier versions were distributed as DbaseIII+ source code, from which later versions were compiled.

9. Notes

1. The utilities CEPH2CRI and CEPH2MAP can be obtained from Steve Bryant, Human Genetic Resources Unit, Imperial Cancer Research Fund, Blanche Lane, South Mimms, Potters Bar, Herts EN6 3LD, England.
2. Software for linkage analysis is regularly discussed in the *Linkage Newsletter*, distributed by Jurg Ott, Columbia University, Box 58, 722 West 168 Street, New York, NY 10032.
3. Under VMS, prepare a file of the following form:

```
$ mapmaker
load-data my.dat
sequence *all
two-point
quit
y
$exit
```

   and give it a .com suffix (e.g., job.com).

   Then submit the job to the batch processor with a command line similar to the following.

```
submit/queue=heavy$queue job.com
```

4. These are quite literally dictionaries of dictionaries, descriptions of and pointers to other collections of data definitions.

References

1. Ott, J. (1985) *Analysis of Human Genetic Linkage.* Johns Hopkins University Press, Baltimore, MD.
2. Elston, R. C. and Stewart, J. (1971) A general model for the genetic analysis of pedigree data. *Hum. Hered.* **21**, 523–542.
3. Morton, N. (1955) Sequential tests for the detection of linkage. *Am J. Hum. Genet.* **7**, 277–318
4. Lange, K. and Elston, R C. (1975) Extensions to Pedigree Analysis I Likelihood calculation for simple and complex pedigrees. *Hum. Hered.* **25**, 95–105

5 Ott, J. (1974) Estimation of the recombination fraction in human pedigrees: Efficient computation of the likelihood for human linkage studies. *Am. J. Hum. Genet.* **26,** 588–597

6. Lathrop, G. M and Lalouel, J. M. (1984) Easy calculations of lod scores and genetic risks on small computers *Am. J Hum. Genet.* **36,** 460–465.

7. Lander, E. S. and Green, P. (1987) Construction of multilocus genetic linkage maps in humans *Proc Natl. Acad. Sci USA* **84,** 2363–2367

8. Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T. P., Bowden, D. W , Smith, D. R., Lander, E. S., Botstein, D., Akots, G., Rediker, K. S , Gravius, T , Brown, V. A , Rising, M. B., Parker, C., Powers, J. A., Watt, D. E., Kauffman, E. R., Bricker, A., Phipps, P., Muller-Kahle, H., Fulton, T. R., Ng, S., Schumm, J W., Braman, J. C , Knowlton, R. G , Barker, D. F., Crooks, S. M., Lincoln, S. E , Daly, M. J., and Abrahamson, J. (1987) A genetic linkage map of the human genome. *Cell* **51,** 319–337.

9 Suarez, B. K. (1978) The affected sib pair IBD distribution for HLA-linked disease susceptibility genes. *Tissue Antigens* **12,** 87–93.

10. Lange, K. (1986) A test statistic for the affected-sib-set method. *Ann. Hum. Genet.* **50,** 283–290.

11. Lange, K. (1986) The affected sib-pair method using identity by state relations. *Am. J. Hum. Genet* **39**(1), 148–150.

12. Weeks, D. E. and Lange, K. (1988) The affected-pedigree-member method of linkage analysis. *Am. J. Hum. Genet.* **42,** 315–326.

13. Karigl, G. (1981) A recursive algorithm for the calculation of identity coefficients. *Ann. Hum. Genet.* **45,** 299–305.

14. Weitkamp, L. R. and Lewis, R. A. (1989) PEDSCORE Analysis of identical by descent (IBD) marker allele distributions in affected family members. *Cytogenet. Cell. Genet.* **51,** 1105–1106.

15 Edwards, A W F (1988) Computers and Genealogies. *Biol. Soc.* **5,** 73–81.

16 MacCluer, J. W , VandeBerg, J. L., Read, B., and Ryder, O. A. (1986) Pedigree analysis by computer simulation *Zoo Biology* **5,** 147–160

17. Boehnke, M (1986) Estimating the power of a proposed linkage study: A practical computer simulation approach *Am. J Hum. Genet.* **39,** 513–527.

18. Ott, J (1989) Computer simulation methods in linkage analysis. *Proc. Natl. Acad. Sci. USA* **86,** 4175–4178.

19. McCarthy, J. L. (1988) The automated data thesaurus: A new tool for scientific information *11th Int. CODATA Conference.* Karlsruhe, Germany.

20. Stephens, J C., Gilna, P., Maglott, D. R , Cavanaugh, M. L., Doute, R. C , Hutchings, G. A., Hayden, J , and Bins, C (1989) Enhancement and expansion of the links between the Gen Bank and ATCC databases in, Human Gene Mapping 10 (1989): 10th Int Workshop on Human Gene Mapping· *Cytogenet. Cell. Genet.* **51,** A2368.

21. Chen, P. P. (1976) The entity-relationship model—toward a unified view of data. *ACM Transactions Database Systems* **1**(1), 9–36.

22 Codd, E. F (1970) A relational model of data for large shared data banks. *Information Retrieval* **13**(6), 377–387

23 Codd, E. F. (1974) Recent investigations in relational data base systems *Information Processing* **74,** 1017–1021.

24. Prokosch H. U , Seuchter S A , Thompson E. A., and Skolnick, M. (1989) Applying expert system techniques to human genetics. *Computers Biomed. Res* **22,** 234–247.

25. Attwood, J. and Bryant, S. (1988) A computer program to make analysis with LIPED and LINKAGE easier to perform and less prone to input errors. *Ann. Hum. Genet* **52**, 259.
26. Mamelka, P. M., Dyke, B., and MacCluer, J. W. (1987) Pedigree/Draw for the Apple Macintosh. Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX.
27. Wolak, G. R. and Sarfarazi, M. (1986) PLOT2000. A pedigree plotting program. Section of Medical Genetics, University Hospital of Wales, Cardiff
28 Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., and Newburg, L. (1987) MAPMAKER· An interactive computer package for constructing genetic linkage maps of experimental and natural populations *Genomics* **1**, 174–181.
29 Lathrop, G. M. and Lalouel, J. M. (1988) Efficient computations in multilocus linkage analysis. *Am. J. Hum Genet.* **42**, 498–505.
30 Lathrop, G. M., Lalouel, J M., Lulier, C., and Ott, J (1984) Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* **81**, 3443–3446.
31. Lathrop, G. M., Lalouel, J. M., Julier, C , and Ott, J. (1985) Multilocus linkage analysis in humans. detection of linkage and estimation of recombination. *Am. J. Hum. Genet.* **37**, 482–498.
32. Hodge, S. E., Morton, L A., Tideman, S., Kidd, K. K., and Spence, M. A. (1979) Age-of-onset correction available for linkage analysis (LIPED). *Am. J. Hum. Genet.* **31**, 761–762.
33. Kernighan, B. and Ritchie, D. (1978) *The C Programming Language*. Prentice-Hall, Englewood Cliffs, NJ.
34. Ploughman, L. M. and Boehnke, M. (1989) Estimating the power of a proposed linkage study for a complex genetic trait. *Am J. Hum. Genet.* **44**(4), 543–551.
35. Wijsman, E. M. (1987) A deductive method of haplotype analysis in pedigrees. *Am. J. Hum. Genet.* **41**, 356–373.
36. Thomas, A. (1987) Pedpack: User's manual. Technical Report No. 99 Department of Statistics, GN-22, University of Washington, Seattle, Washington 98195.
37. Thomas, A. (1987) Pedpack. Manager's manual. Technical Report No. 100. Department of Statistics, GN-22, University of Washington, Seattle, Washington 98195.
38. Hasstedt, S. J. and Cartwright, P. E. (1981) PAP—Pedigree Analysis Package University of Utah, Department of Medical Biophysics and Computing, Technical Report No 13. Salt Lake City, UT.
39. Lange, K., Weeks, D., and Boehnke, M. (1988) Programs for Pedigree Analysis—Mendel, Fisher and Dgene. *Genet. Epidem.* **5**(6), 471,472
40. Morton, N. E. and Andrews, V (1989) MAP, An expert system for multiple pairwise linkage analysis. *Ann. Hum. Genet.* **53**, 263–269.

CHAPTER 32

Creating Animal Models of Genetic Disease

Robert P. Erickson

1. Introduction

There are many methods for creating animal models of human genetic disease, and it is not possible for a short chapter to provide complete details of the methods used. There are four general approaches that will be discussed. One involves embryonic stem cells and another transgenic mice, each of which are the subject of major portions of other methods books *(1,2)*. This chapter will provide only an overview of the general approaches and some specific details about several of them. Additionally, it provides references to primary papers or other review sources, in order to enable the investigator to find further details of the methods. Each of the methods has particular strengths and weaknesses, and a goal of this chapter is to help readers choose the approach that might be appropriate to their problem.

An important first step, before setting out to create an animal model of genetic disease, is to see if an appropriate one already exists. The literature on this subject is frequently not well known by many people interested in human disease, since much of it appears in the veterinary or other specific literature. Thus, Table 1 is provided as a source of some review articles and books that describe many animal models of genetic disease. It is important to realize that the species involved vary greatly—from cattle to mice. Although α-mannoidosis or citrullinemia may be economically important in cattle, and therefore of great interest to veterinarians, most individuals will not be able to afford a herd of cattle for experimental purposes! Dogs, cats, rabbits, and

From. *Methods in Molecular Biology, Vol 9 Protocols in Human Molecular Genetics*
Edited by: C Mathew Copyright © 1991 The Humana Press Inc , Clifton, NJ

Table 1
Useful Reviews of Currently Available Animal Models

Species	Topic or subtopic	Reference
Mice	General	*4*
	Birth defects	*5*
	Anemias	*67*
	Neurological	*68*
		69
Various	General	*3, 70, 71*
	Connective tissue	*72*
	Hemorrhagic and thrombotic	*73*
	Glycogen storage disease	*74*
	Retinal degeneration	*75*

miniature swine are all animals of a size amenable to many procedures used in humans and, therefore, are sought-after animal models. Since most of the methods for creating animals models of genetic disease involve mice, I would like to briefly allay fears about difficulties in using mice for medical research. Although mice are the mammalian genetic model par excellence because of their rapid breeding time and relatively small size, many physiologists have been prejudiced against using them. However, given current advances in miniaturization of sensors, catheters, and so on, most procedures that can be performed on humans, cats, or dogs can also be performed on mice. Thus, there should be no inherent prejudice against using mice as animal models. Of the many references listed in Table 1, one in particular should be singled out as a general compendium. This is *Handbook: Animal Models of Human Diseases,* which is a continuing compilation of animal models from regular articles appearing in the *American Journal of Pathology* edited by Copen, Jones, and Migaki *(3)*. This could be a starting point for many individuals. Other important references are the reviews of already existing mouse models by Leiter et al. *(4)* and Kalter *(5)*.

1.1. Strategy

Of the four methods to be described—mutations with directed screening, transgenic mice, homologous replacement or mutation in embryonic stem cells, and antisense techniques—each has special requirements and special capabilities, which will be discussed briefly here, and each of the methods will be discussed more fully below. Some of the successes with the various methods are described in Table 2.

Table 2
Some Created Animal Models of Genetic Disease

Disease/mutation	How made	Ref.
β-Thalassemia	Spontaneous mutation, but found in control for mutation screening	*76*
Hemoglobin Rainier	Directed screening after ENU mutagenesis	*77*
Hyperphenylalaninemia	Directed screening after ENU mutagenesis	*8*
Muscular dystrophy	Directed screening after ENU mutagenesis	*9*
Carbonic anhydrase II deficiency	Directed screening after ENU mutagenesis	*78*
α_1-Antitrypsin deficiency	Human α_1-antitrypsin A allele in transgenic mice	*22* *23*
Osteogenesis imperfecta	Mutated collagen gene in transgenic mice	*28*
Type I diabetes mellitus	Major histocampatability complex antigens—and interferon—expression in transgenic mice	*30* *31* *29*
Microphthalmia	Ablating lens cells in transgenics for crystallin promoter/toxin construct	*36* *37*
Pituitary dwarfs	Ablating growth hormone expressing cell in transgenics for growth hormone promoter/toxin construct	*38*
$HPRT^-$ variant deficiency	Spontaneous $HPRT^-$ variant of embryonic stem cells introduced into mice	*41*

The method of mutation and directed screening does not require a cloned gene, but does require a clearly identified phenotype. Thus, if one has located a gene product, readily detectable by electrophoresis, that one wishes to alter or eliminate, mutation and directed screening may be the most desirable approach. This approach may also be used to create morphological variants, although the criteria for the correct mutation are usually much "softer." The screening usually requires a large number of animals and efficient procedures.

The use of transgenic mice requires special technical capabilities and a cloned gene. However, it requires a smaller number of mice and can be the approach of choice. It has been used to create animal models by over-producing gene products, producing abnormal gene products in the presence of normal gene products, expressing normal genes at inappropriate times or places, and by ablating certain cell types.

Homologous gene replacement in embryonic stem cells offers the potential of creating a desired mutation in cultured cells. It, too, requires a

cloned gene and special embryo-handling techniques. Although such homologous replacements have been achieved multiple times, to date it has been difficult to get the resulting cells to contribute to the germ line of the animals resulting from the embryos in which they have been placed. Although several approaches to maintaining the potential to contribute to the germ line are being studied, this is still a drawback with this method.

Finally, the use of antisense techniques offers the opportunity to create conditional, rather than merely null, mutations when one has a cloned gene, but this approach has not yet been very successful in creating animal models of human disease. The one success to date replicates a disease already known in mice, but not yet known in humans.

2. Methods

2.1. Mutation and Directed Screening

The availability of mutagenic agents, such as ethylnitrosourea (ENU), which can generate mutations at frequencies on the order of one in 2000 per gamete per generation *(6)*, makes it now practical to deliberately induce mutations. A clear-cut phenotype is required. The largest amount of experience has been gained in searching for mutations with electrophoretic techniques. One particularly useful method of screening for electrophoretic mutations has been developed by Johnson and Lewis *(7)*. It takes advantage of differences in electrophoretic mobility for a number of enzymes between two different inbred strains of mice. If a male or female of one of the inbred strains is treated with a mutagen and mated to the other inbred strain, offspring with a null mutation can be detected because only the nonmutated parental electrophoretic band will appear, instead of the double (or more complex) banding pattern characteristic of the F1 hybrid. For those enzymes for which there are not electrophoretic variations between the two inbred strains, mutations that alter the charge of an enzyme or protein are detectable, but null mutations, which would merely decrease the signal by one-half, are not usually detectable. Such induced-mutation methods can also be used to screen for dominant visibles or for phentoypes other than those detectable by electrophoresis. If one wishes to detect a recessive mutation, however, the possibly mutated progeny have to be mated and the resulting offspring mated back to the parent. This adds two generations to the screening procedure. Nonetheless, this has been a successful method to detect hyperphenylalaninemic mice by screening for elevated phenylalanine in the urine by the standard test used for human screening for phenylketonuria *(8)*. In the case of X-linked mutations, male progeny can show the mutation directly, and this approach has been used to generate extra mutations at the mouse equivalent of the human Erbe-Duchenne muscular dystrophy (dystrophin) locus, ***mdx*** *(9)*.

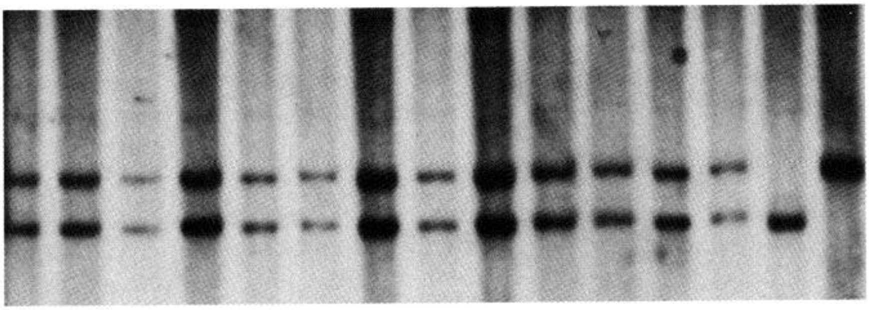


Fig. 1. Screening for deletions of c-*met* in offspring of irradiated mice. Lanes 1–13, individual F_1 mice; Lane 14, parental C57BL6/J; Lane 15, parental DBA/2J.

In the case in which one has a marker near a gene, but not a marker for the gene itself, it should be possible to create radiation-induced deletions that would delete both the marker and the gene of interest. We have attempted to use this method to create a mouse model of cystic fibrosis. The c-*met* oncogene is closely linked to cystic fibrosis in humans, and it is likely to be linked to a mouse equivalent of cystic fibrosis, since several loci on human chromosome 7 remain linked together on mouse chromosome 6 *(10)*. We found an electrophoretic difference in the size of hybridizing DNA, after restriction enzyme digestion and Southern transfer, between the DBA/J and C57BL/6J inbred strains with the c-met oncogene, i.e., a restriction fragment length polymorphism, or RFLP *(11)*. We screened a large number of off-spring from irradiated parents but all of the F1 mice showed both parental bands—we were not able to detect a deletion removing the parental c-*met* allele in over 1400 animals screened (Fig. 1). However, the particular dose of radiation used was not creating as many deletions as expected, so our failure does not indicate that this is not a potentially useful approach. Chlorambucil has recently been found to induce deletions and/or translocations at a very high frequency in spermatids *(12)*, and it can be used in a similar manner.

ENU is a hazardous substance, and there are several technical difficulties involved in using it. It may be purchased in 2-g lots in individual vials (± 0.1 g) from the Radian Corporation, Austin, TX. Since ENU is quite unstable, it needs to be stabilized in acid. Each vial of ENU can be dissolved in the appropriate volume of a pH-5 buffer containing 24.3 m*M* citric acid and 51.4 m*M* sodium phosphate. ENU-containing buffer solutions are constituted in such a manner that a 0.3-mL intraperitoneal injection will deliver the desired dosage (mg/kg) of ENU. Doses from 100 to 250 mg/kg have been used. One can choose to give ENU to females, in which case various stages of

meiosis will be sampled. Alternatively, if males are used, there will be a presterile period in which the gametes from late stages of spermatogenesis have been exposed to ENU. After recovery from the sterile period, the gametes will result from spermatogonial stages that have been exposed to ENU. Since ENU is highly toxic and special chemical precautions must be used when handling it, this would not be a practical approach in many laboratory settings.

Another approach to generating the mutations for directed screening, one that allows ready cloning of the gene, is insertional mutagenesis with retroviruses, or transgene injection *(13)*. However, the efficiencies of mutation achieved by insertional mutagenesis are sufficiently low that directed screening has not yet been performed, although interesting mutants have been found *(14)*.

2.2. Transgenic Mice

The introduction of transgenic-mouse technology by Gordon and Ruddle *(15,16)* was a major advance in biotechnology. The technique has become the reference standard for studying gene regulation in mammals and has allowed testing "Koch's postulates" for a genetic disorder: Transgenic replacement of growth hormone, in dwarf "little" *(lit)* mice (which were thought to be dwarfed because of growth-hormone deficiency), corrected their growth and their decreased fertility *(17)*. To the extent that cancer is a genetic problem, one could consider those transgenic mice in which oncogenes are activated in particular tissues, and which develop early and severe carcinomas, as models of genetic disease *(18,19)*. Another not-quite-genetic animal model created in transgenic mice is that for progressive multifocal leukoencephalopathy, which was created by making mice transgenic for the early region of human papovavirus JC—some of these mice exhibited dysmyelinization in the central nervous system, comparable to this disorder *(20)*.

One class of animal models created in transgenic mice is that caused by overproduction of a protein coded by the transgene. Sasaki et al. *(21)* created a possible mouse model for familial amyloidotic polyneuropathy by making transgenic mice for a human transthyretin variant. A particular form of familial amyloidotic polyneuropathy is associated with a methionine-for-valine substitution at position 30 in plasma transthyretin. It is believed that this substitution results in the systemic amyloidosis, with prominent peripheral nerve involvement causing the symptomatology. These researchers used the mouse metallothionein-I promoter to express the human transthyretin variant in the serum of transgenic mice. Another mouse model of genetic disease created by making transgenic mice that overproduce a gene product is that for α_1–antitrypsin deficiency. Although mice produce their own α_1–antitryp-

sin, it was found that mice overproducing the human Z-allele of α_1–antitrypsin accumulated this form of the molecule in cytoplasmic droplets and developed a liver pathology comparable to that seen in some humans that are homozygous for this allele *(22,23)*. However, these mice, which maintain their endogenous α_1–antitrypsin, have not yet been found to have any pulmonary pathology, which is the nearly invariant feature of the human deficiency. Nomura et al. *(24)* have created transgenic mice overproducing renin in the hope of creating animal models with hypertension. There are limitations to the transgenic approach; inclusion of the dominant control region of the human β-globin locus *(25)* in a β^S-globin (β chain or Hemoglobin S [HbS] causing sickle-cell anemia) transgene led to the production of a transgenic mouse in which HbS constituted 83% of the total hemoglobin *(26)*. However, the mouse showed no obvious manifestations of homozygous sickle-cell disease.

Dominant mutations must ultimately be explained by altered genetic regulation or some aspect of negative complementation. Herskowitz *(27)* has emphasized the potential for producing dominant mutations at will, either by such negative complementation or by overproduction of an abnormal protein leading to abnormal function even in the presence of the normal protein. There is one outstanding example of the successful use of the gene for a mutant protein to cause a disorder in transgenic mice. Stacey et al. *(28)* performed in vitro mutagenesis on α-1 collagen, introducing mutations similar to those causing osteogenesis imperfecta in humans. When introduced into transgenic mice, a marked similarity in pathology to that found in perinatal lethal osteogenesis imperfecta (OI II) was found. These experiments allowed the exploration of quantitative relationships between the abnormal collagen and the disease state—an amount as low as 10% of normal of the abnormal collagen caused a severe disorder despite the continued presence of normal amounts of the normal collagen.

Although the cancerous mice discussed above sometimes result from expression of normal genes (c-oncogenes) in ectopic locations or at abnormal times, the clearest animal model of an at least partially genetic disease resulting from the expression of normal genes at inappropriate times or places in transgenic mice has been the production of type I diabetes mellitus in mice. The pancreatic islet, β-cell-specific insulin promoter was used to target expression of genes for class histocompatibility antigens *(29)*, class II histocompatibility antigens *(30,31)*, and interferon γ *(30)*, which can increase expression of major histocompatibility (MHC) antigens. These genes were expressed in β-cells and resulted in the development of insulin-dependent diabetes in the mice. Interestingly, evidence of T-cell infiltration and autoimmunity to these β-cells was not found in the mice expressing the major histo-

compatibility antigens, but was found in the ones expressing interferon γ. This suggests a nonimmune role for the transgenic major histocompatibility molecules in the impairment of β-cell function, which might be related to the role of MHC molecules in hormone expression *(32)*.

An alternative approach utilizing transgenic mice involves using developmental and/or tissue-specific promoters to express toxic genes capable of killing particular cell lineages. The strategy has been named "toxigenics" *(33)*. Although originally performed with intracellular toxins such as ricin and diphtheria toxin-A polypeptide *(34)*, an alternative approach took advantage of a viral enzyme that can utilize the antiherpetic drug gancyclovir, killing the cells with the toxic metabolite. Thus, tissue specificity could be achieved by varying the promoter for the herpes thymidine kinase gene and the time of ablation could be chosen by timing the administration of the drug *(35)*. The toxigenic method has been used by two groups to create microphthalmic mice. In both cases, lens cells were ablated in transgenic mice by using a crystallin protein promoter to drive the toxin gene *(36,37)*. In addition, Behringer et al. *(38)* have created pituitary dwarfs by eliminating cells that express growth hormone in the pituitary in transgenic mice expressing a toxin driven by the growth hormone promoter.

Creating transgenic mice requires fairly sophisticated equipment, which is well described in the methods manual of Hogan et al. *(2)*. Skilled operators perform pronuclear injections at the rate of about 100 embryos/hour, of which at least half should survive the injection procedure. DNA has to be highly purified for the injection, and the preparation methods are also described in this manual. There are many variables in the choice of promoters and design of constructs for injection. Extensive results by the Brinster–Palmiter collaboration, who have found that constructs containing an intron are usually much better expressed, are summarized in an important paper *(39)*. An excellent way to start using this technology is to take Cold Spring Harbor Laboratory's intensive, "hands-on" course in manipulating the mammalian embryo.

2.3. Embryonic Stem Cells

Embryonic stem (ES) cells are totipotent cells that can be cultured from early embryos *(40)*. They can be maintained and manipulated in culture and subsequently introduced into blastocysts, sometimes contributing to the germ line. Thus, mutants selected in vitro or genes inserted into ES cells can be introduced into the mouse germ line. Two groups have successfully selected ES cells with defects in hypoxanthine guanine phosphoribosyl transferase (HPRTase) and have reintroduced the stem cells into mouse zygotes *(41,42)*. The original chimeras were bred, and eventually homozygous HPRTase-deficient mice were developed. However, these mice have apparently not devel-

oped some of the classical symptoms of Lesch-Nyhan disease, such as self-mutilatory behavior or symptoms of gout *(43)*. It is possible that mice are protected from potentially toxic metabolites of uric acid, inasmuch as mice, but not humans, can convert uric acid to allantoin by oxidation with urate oxidase. Thus, it may be necessary to eliminate both HPRT activity and urate oxidase to succeed in creating an animal model of Lesch-Nyhan disease. Since the gene for urate oxidase has been cloned *(44)*, ablation of this enzyme for an animal model should be possible. However, there are only a few genes for which a deficiency creates the situation in which in vitro selection can be performed to isolate the mutant-bearing cell, so other techniques are needed.

Replacement of a gene in ES cells by a specifically altered form through the mechanism of homologous recombination offers an alternative method for creating animal models of genetic disease (*see also* Chapter 20, this volume). Inasmuch as the insertion site is the normal gene, no artifacts attributable to the insertion site (or to chance of insertional mutagenesis) should occur. This may also be the ideal method for correcting defective genes in bone-marrow stem cells (or in other cell types) for gene therapy. Smithies et al. *(45)* first demonstrated that transfected cloned genomic sequences would recombine homologously with the endogenous gene at a frequency of about 1/1000. Capecchi and coworkers have studied the conditions affecting homologous recombination into the single HPRTase gene of a male ES cell line *(46)*. A neomycin-resistance marker was used, the insertion of which would inactivate the HPRTase gene, and which allowed selection for transfected cells with G418. In addition, the inactivation of HPRTase made the cells resistant to 6–thioguanine, providing a rapid selection system for homologous recombination. When sufficient flanking homologous HPRTase sequences were present, 1/1000 of the neomycin-resistant colonies were the result of homologous recombination.

Recently Smithies and colleagues have corrected a previously mutated HPRT gene by homologous replacement in ES cells *(47)*, and Thompson et al. *(48)* were successful in generating a line of mice from ES cells in which a mutant HPRT gene has been corrected by this technique. Several groups have used the basic notion of letting the promoter of a target gene activate the selectable marker (which would otherwise not be expressed) for selection of homologous replacement events *(49–51)*. Of course, this method requires that the target gene be one that is expressed in the ES cell line. Most recently Capecchi's group *(52)* have introduced a combined positive- and negative-expression system to select for homologous replacement events. In this approach, both a neomycin-resistance gene and a herpes simplex thymidine kinase gene are used in the construct. Random integration of the construct will lead to both neomycin resistance and, because of the cointegration of the thymidine kinase gene, gancyclovir sensitivity. However, with homolo-

gous recombination, the neomycin-resistance gene, which is centrally positioned in the construct, will be incorporated, but the herpes simplex virus thymidine kinase, which is at an end of the construct and would not be integrated by homologous recombination, will be lost, thus making the cell line resistant to gancyclovir. Capecchi's group have disrupted the protooncogene *int-2* in mouse ES cells, but have not yet created a mouse line bearing the mutation *(52)*.

The methods used in deriving ES cells from preimplantation mouse embryos are well described in an article by Robertson in the book that she has edited *(1)*. ES cells are usually grown on feeder cells, which introduces moderate complexity to the tissue-culture methods. Recently it has been shown that leukemia inhibitory factor (LIF) will substitute for this requirement *(53,54)*. Although ES cell lines can be obtained from other investigators, a general prejudice is that it is better to derive one's own—the longer they have been maintained in culture, the greater the likelihood of chromosomally visible, i.e., karyotypic, changes, and potentially, invisible changes that prevent the ES cell lines from contributing to germ-line chimeras. It is possible that LIF could help prevent such changes. Thus, the technology for working with ES cells is rapidly evolving. The methodology for injecting them into embryos is covered in both the Hogan et al. *(2)* and Robertson *(1)* methods books. The technique requires micromanipulation equipment generally similar to that needed to make transgenic mice. Obtaining such equipment may be a major obstacle, preventing many groups from using this approach to creating animal models of genetic disease. Also, given the current low rate of contribution of mutagen-treated ES cells to the germ line, it might be wise to wait for improved methods of generating germ-line chimeras before investing heavily in this approach.

With appropriate constructs it should be possible to identify the homologous recombinants by the polymerase chain reaction (PCR) and not need to depend on selection (Fig. 2). Such constructs would not necessarily need the neomycin-resistance marker, since transfection can be as much as 40% efficient *(55)*. The major feature of the vector would be the inclusion of a PCR oligomer-primer binding site in correct orientation to a second primer site outside the target gene—then a PCR fragment of known size would occur only with homologous recombination. Pools of transfected cells would allow detection of the pool with a homologous recombinant; subdividing the pool would lead to identification of the desired cell line for placement in blastocysts and for generation of mouse chimeras.

2.4. Antisense RNA

Antisense RNA is a mode of genetic regulation used naturally by prokaryotes *(56)* that is more recently being applied experimentally in eukaryotes

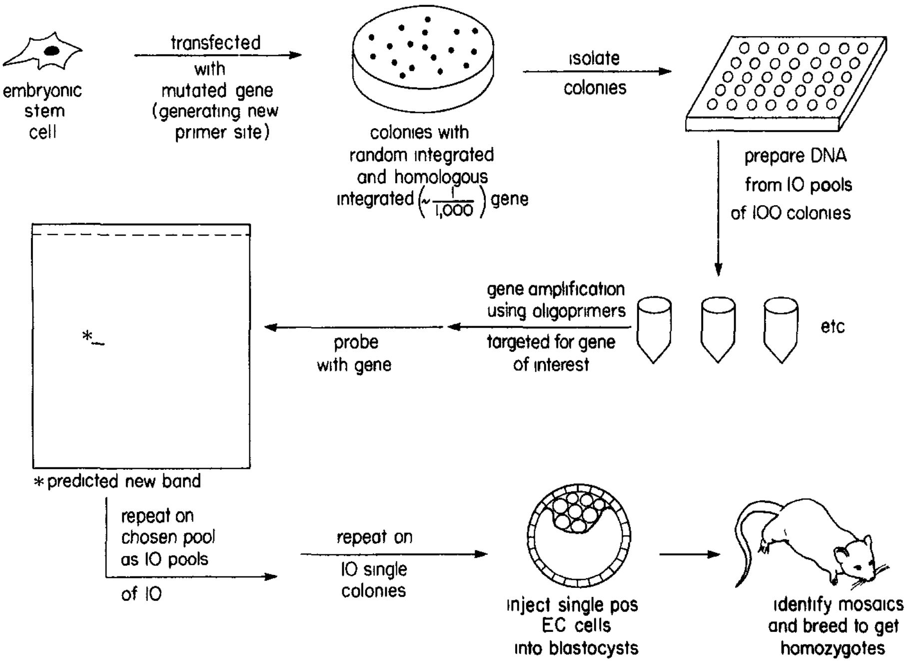


Fig. 2. A generalized schema for creating directed mutations in mice by using homologous recombination in embryonic stem cells. *See text* for details. From *Am. J. Hum. Genet.* **43**, 584, by permission.

(57). Either antisense RNA or antisense DNA (either alone or in constructs expressing antisense RNA) has been used to inhibit the expression of a wide variety of cloned genes and to study their function (reviewed in ref. *56*). More recently, this technique has found application in the genetic analysis of development in *Drosophila (58,59)* and *Xenopus (60)*. We have been utilizing the antisense RNA approach to study the functions of genes expressed during preimplantation development in the mouse and have demonstrated its applicability to preimplantation mouse embryos *(61–63)*.

A genetic disease has been created in mice transgenic for an antisense construct for deficiency of myelin basic protein *(64)*. This deficiency is not a model for a human genetic disease, since no human disorder is known to be caused by a defect in myelin basic protein. Nonetheless, the successful use by this group of antisense methodology to partially eliminate the function of an

important gene shows the practicality of the approach. A deficiency of myelin basic protein causes a disorder called *shiverer* in mice, the phenotype of which is described by the name. In the antisense work, the authors used the myelin basic protein promoter to direct antisense mRNA synthesis. A reduction of 80–90% of myelin basic protein messenger RNA levels was achieved. However, symptoms occurred only when the antisense transgene was placed in heterozygotes for the *shiverer* gene. Thus, if a disease requires a very marked deficiency of protein product, then it is currently less practical to achieve its inhibition with the antisense approach. Nonetheless, as more promoters are studied, higher levels of inhibition may be achieved. Also, there are a number of diseases that may appear when there are smaller reductions of the gene product. For instance, nuclear-encoded mitochondrial subunits may be so essential that even partial reductions would lead to a significant disorder.

The promoter for heat-shock protein 68 (hsp 68) should be useful for antisense work. As shown by Rossant's group (Kathory et al., ref. *65*) this promoter (–664 to +113) is excellently and specifically inducible by short-term treatments with heat or with sodium arsenate in preimplantation and gestational stages of development. It should similarly be controllable postnatally. Thus, mouse cDNAs can be cloned in antisense orientation to this promoter. With this promoter, it seems unlikely that there would be any prenatal lethality without induction and, thus, pronuclear-injected embryos can no doubt be transferred to pseudopregnant females. After birth, DNA can be prepared from tail tips and animals screened for the presence of the transgene. When liveborn offspring are obtained with the transgene, these can be studied pathologically. Positive transgenic lines can be established, creating the situation where half the progeny of an animal will carry the transgene. In this case, one can take groups of pups and treat them with sodium arsenate by injection and/or brief periods of heat exposure and study the pathological phenotype. In addition, since there should not be any activity of this promoter prior to induction, one can breed animals homozygous for the transgene. This would increase the levels of antisense obtained after the induction. In all cases, animals should be characterized with DNA prepared from their tail tips either by direct Southern blotting or by PCR assays. The PCR assays take advantage of unique fragments that would come from the clone (since the construct brings together mouse genes not normally in proximity to each other).

It should be possible to inhibit genes in particular tissues using tissue-specific promoters. For instance, several groups have clearly shown that the human cardiac actin gene promoter is highly muscle-specific and leads to high-level transcription of this gene in differentiated muscle cells *(66)*. Again, one can clone cDNAs in reverse (and forward) orientation to this promoter.

Thus, a variety of promoters are available that should allow careful design of tissue-, time-, and developmental-stage-specific activation of an antisense gene. However, as already discussed, the success of this approach will depend on incomplete inactivation of the gene creating the disease state.

3. Conclusion

This chapter has discussed a number of approaches to creating animal models of genetic disease. They are all designed to be used with mice, but could be used in other species as veterinarians gain more experience with the transgenic approach in domestic species. However, for reasons of cost and convenience, most investigators would want to use mice. Although a number of examples of successful creation of animal models of genetic disease have been given, it is not yet fully apparent how useful they will be for medical research. Many of them have been found so recently that they have not yet been extensively used for investigations. Nonetheless, exciting recent developments, such as the cloning of the cystic fibrosis gene, will certainly lead to more animal models of genetic disease being created by some of the approaches discussed in this chapter.

Acknowledgments

I thank Stan R. Blecher for useful comments, Judy Worley for secretarial assistance, and the Cystic Fibrosis Foundation for research support.

References

1. Robertson, E J (1986) *Teratocarcinomas and Embroyonic Stem Cells: A Practical Approach* IRL, Oxford, UK and Washington, DC. pp. 1–254.
2. Hogan, B., Costantini, F., and Lacy, E (1986) *Manipulating the Mouse Embryo: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. pp. 1–332.
3. Copen, C. C , Jones, T. C , and Migaki, G. (1986) *Handbook Animal Models of Human Disease,* 15th fascile. Registry of Comparative Pathology, Armed Forces Institute of Pathology, Washington, DC.
4. Leiter, E H , Beamer, W G , Shultz, L. D., Barker, J. E., and Lane, P. W. (1987) Mouse models of genetic disease. *Birth Defects* **23,** 221–257.
5. Kalter, H. (1980) A compendium of the genetically induced congenital malformations of the house mouse *Teratology* **21,** 397–429.
6. Johnson, F M and Lewis, S E (1981) Electrophoretically detected germinal mutations induced by ethylnitrosourea in the mouse *Proc. Natl Acad Sci. USA* **78,** 3138–3141.
7. Johnson, F. M. and Lewis, S E (1982) The human genetic risk of airborne genotoxics. An approach based on electrophoretic techniques applied to mice, in *Genotoxic Effects of Airborne Agents* (Tice, R , Costa, A , and Schaido, K., eds.), Plenum, New York, pp 596–606

8. McDonald, J. D. and Bode, V. C. (1988) Hyperphenylalaninemia in the *hph-1* mouse mutant. *Pediatr. Res.* **23,** 63–67.
9. Chapman, V. M., Miller, D. R., Armstrong, D., and Caskey, C. T. (1989) Recovery of induced mutations for X-chromosome-linked muscular dystrophy in mice. *Proc. Natl. Acad. Sci. USA* **86,** 1292–1296.
10. Nadeau, J. H. (1989) Maps of linkage and synteny homologies between mouse and man. *Trends Genet.* **5,** 82–86.
11. Sweet, A. M., Cohen, J. A., Lopez, M., and Erickson, R. P. (1988) An EcoRI polymorphism for pmetG in mice. *Nucleic Acids Res.* **16,** 8745.
12. Russell, L. B., Hunsicker, P. R., Cacheiro, U. L. A., Banghorn, J. W., Russell, W. L., and Shelby, M. D. (1989) Chlorambucil effectively induces deletion mutations in mouse germ cells. *Proc. Natl. Acad. Sci. USA* **86,** 3704–3708.
13. Gridley, T., Soriano, P., and Jaenisch, R. (1987) Insertional mutagenesis in mice. *Trends Genet.* **3,** 162–166.
14. Woychick, R. P., Stewart, T. A., Davis, L. G., D'Eustachio, P., and Leder, P. (1985) An inherited limb deformity created by insertion mutagenesis in a transgenic mouse. *Nature* **318,** 36–40
15. Gordon, J. W., Scangos, G. A., Plofkin, D. J., Barbosa, J. H., and Ruddle, F. H. (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc. Natl. Acad. Sci. USA* **77,** 7380–7384.
16. Gordon, J. W. and Ruddle, F. H. (1981) Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* **214,** 1244–1246
17. Hammer, R. E., Palmiter, R. D., and Brinster, R. L. (1984) Partial correction of murine hereditary growth disorder by germ-line incorporation of a new gene. *Nature* **311,** 66–67
18. Stewart, T. A., Pattengale, P. K., and Leder, P. (1984) Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/*myc* fusion genes. *Cell* **38,** 627–637.
19. Brinster, R. L., Chen, H. Y., Messing, A., van Dyke, T., Levine, A. J., and Palmiter, R. D. (1984) Transgenic mice harboring SV40–antigen genes develop characteristic brain tumors *Cell* **37,** 367–379.
20. Small, J. A., Scangos, G. A., Cork, L., Jay, G., and Khoury, G. (1986) The early region of human papovavirus JC induces dysmyelination in transgenic mice. *Cell* **46,** 13–18.
21. Sasaki, H., Tone, S., Makazato, M., Yoshioka, K., Matsuo, H., Kato, Y., and Sakaki, Y. (1986) Generation of transgenic mice producing a human transthyretin variant: A possible mouse model for familial amyloidotic polyneuropathy. *Biochem. Biophys. Res. Comm.* **139,** 794–799
22. Dycaico, M. J., Grant, S. G. N., Felts, K., Nichols, W. S., Geller, S. A., Hagger, J. H., Pollard, A. J., Kohler, S. W., Short, H. P., Jirik, F. R., Hanahan, D., and Sorge, J. A. (1988) Neonatal hepatitis induced by α_1–antitrypsin in a transgenic mouse model. *Science* **242,** 1409–1412.
23. Carlson, J. A., Rogers, B. B., Sifers, R. N., Finegold, M. J., Clift, S. M., DeMayo, F. J., Bullock, D. W., and Woo, S. L. C. (1989) Accumulation of PiZ α_1-antitrypsin causes liver damage in transgenic mice. *J. Clin. Invest.* **83,** 1183–1190.
24. Nomura, T., Katsuki, M., Yokoyama, M., and Tajima, Y. (1987) Future perspectives in the development of new animal models, in *Animal Models: Assessing the Scope of Their Use in Biomedical Research* (Kawamata, J. and Melby, Jr., E. C., eds), Alan R. Liss, NY, pp. 337–353.

25. Grosveld, F , van Assendelft, B., Greaves, D. R., and Kollias, G. (1987) Position independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51**, 976–985.
26. Greaves D. R., Fraser P., Vidal M. A., Hedges M. J., Ropers D., Luzzatto L., and Grosveld F. (1990) A transgenic mouse model of sickle cell disorder. *Nature* **343**, 183–185.
27. Herskowitz, I. (1987) Functional inactivation of genes by dominant negative mutations. *Nature* **329**, 219–222.
28 Stacey, A., Bateman, J., Choi, T., Mascara, T., Cole, W., and Jaenisch, R. (1988) Perinatal lethal osteogenesis imperfecta in transgenic mice bearing an engineered pro-α_1(I) collagen gene. *Nature* **332**, 131–136.
29. Allison, J., Campbell, I. L., Morahan, G., Mandel, T. E., Harrison, L. C., and Miller, J. F. (1988) Diabetes in transgenic mice resulting from over-expression of class I histocompatibility molecules in pancreatic β-cells. *Nature* **333**, 529–533.
30. Sarvetnick, N., Liggitt, D., Pitts, S. L., Hansen, S. E., and Stewart, T. A. (1988) Insulin-dependent diabetes mellitus induced in transgenic mice by ectopic expression of class II MHC and interferon-gamma. *Cell* **52**, 773–782.
31. Lo, D., Burkly, L. C., Widera, G., Cowing, C., Flarell, R. A , Palmiter, R. D., and Brinster, R L. (1988) Diabetes and tolerance in transgenic mice expressing class II MHC molecules in pancreatic beta cells. *Cell* **53**, 159–168.
32. Erickson, R. P. (1986) Glucagon receptor number and the MHC. *Nature* **323**, 586.
33. Beddington, R. S. P. (1988) Toxigenics· Strategic cell death in the embryo. *Trends Genet.* **4**, 1–2.
34. Palmiter, R D., Behringer, R. R., Quaife, C. J., Maxwell, F , Maxwell, I. H., and Brinster, R. L. (1987) Cell lineage ablation in transgenic mice by cell-specific ablation of a toxin gene. *Cell* **50**, 436–443.
35. Heyman, R. A., Borrelli, E , Lesley, J , Anderson, D., Richman, D. D., Bavrd, S. M., Hyman, R., and Evans, R. M. (1989) Thymidine kinase obliteration: Creation of transgenic mice with controlled immune deficiency. *Proc. Natl. Acad. Sci. USA* **86**, 2698–2702.
36. Breitman, M. L., Clapoff, S , Rossant, J., Tsui, L.-C., Golden, L. M., Maxwell, I. H., and Bernstein, A. (1987) Genetic ablation: Targeted expression of a toxin gene causes micropthalmia in transgenic mice *Science* **238**, 1563–1565.
37. Landel, C. P., Zhao, J , Bok, D., and Evans, G. A. (1988) Gene specific expression of recombinant ricin induces developmental defects in the eyes of transgenic mice. *Genes Dev.* **2**, 1168–1178.
38. Behringer, R. R., Mathews, L S , Palmiter, R. D., and Brinster, R. L (1988) Dwarf mice produced by genetic ablation of growth hormone-expressing cells. *Genes Dev.* **2**, 453–461.
39 Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E., and Palmiter, R. D. (1988) Introns increase transcriptional efficiency in transgenic mice. *Proc. Natl Acad. Sci. USA* **85**, 836–840.
40. Evans, M. J. and Kaufman, M. H. (1981) Establishment in culture of pluripotential cells from mouse embryos *Nature* **292**, 154–156.
41. Hooper, M., Hardy, K., Handyside, A , Hunter, S , and Monk, M (1987) HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germ line colonization by cultured cells. *Nature* **326**, 292–295.
42 Kuehn, M. R., Bradley, A , Robertson, E. J., and Evans, M. J. (1987) A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. *Nature* **326**, 296–298.

43. Stout, J. T. and Caskey, C. T. (1988) The Lesch-Nyhan syndrome: Clinical, molecular, and genetic aspects. *Trends Genet.* **4,** 176–178.
44. Lee, C. C., Wu, X., Gibbs, R. A., Cook, R. G., Muzny, D. M., and Caskey, C. T. (1988) Generation of cDNA probes directed by amino acid sequence: Cloning of urate oxidase. *Science* **239,** 1288–1290.
45. Smithies, O., Gregg, R. G., Boggs, S. S , Koralewski, M. A., and Kucherlapati, R. S. (1985) Insertion of DNA sequences into the human chromosome β-globin locus by homologous recombination. *Nature* **317,** 230–234
46. Thomas, K. R. and Capecchi, M. R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51,** 503–512.
47. Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S., and Smithies, O. (1987) Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330,** 576–578.
48. Thompson, S., Clarke, A. R., Pow, A. M., Hooper, D , Melton, D. W. (1989) Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* **56,** 313–321.
49. Sedivy, J. M. and Sharp, P. A. (1989) Positive genetic selection for gene disruption in mammalian cells by homologous recombination *Proc. Natl. Acad. Sci. USA.* **86,** 227–231.
50. Jasin, M. and Berg, P. (1988) Homologous integration in mammalian cells without target gene selection. *Genes Dev.* **2,** 1353–1363.
51. Doring, J. R., Inglis, J. D., and Porteous, D. J. (1989) Selection for precise chromosomal targeting of a dominant marker by homologous recombination. *Science* **243,** 1357–1360.
52. Mansour, S. L., Thomas, K. R , and Capecchi, M. R (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: A general strategy for targeting mutations to non-selectable genes. *Nature* **336,** 348–352.
53. Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., and Rogers, D. (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336,** 688–690.
54. Williams, R. L., Hilton, D. J., Pease, S., Wilson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A., and Gough, N. M. (1988) Myeloid leukanemia inhibitory factor maintains the developmental potential of embryonic stem cells *Nature* **336,** 684–687.
55. Chen, C. and Okayama, H. (1987) High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.* **7,** 2746–2752.
56. Green, J., Pines, O , and Inouye, M. (1986) The role of antisense RNA in gene regulation. *Annu. Rev. Biochem.* **55,** 569–597.
57. Izant, J. G. and Weintraub, H (1985) Constitutive and conditional suppression of exogenous and endogenous genes by antisense RNA. *Science* **229,** 346–352
58. Rosenberg, U. B., Priess, A , Seifert, E., Jackle, H , and Knipple, D. C. (1985) Production of phenocopies by *Kruppel* antisense RNA injection into *Drosophila* embryos *Nature* **313,** 703–706.
59 Cabrera, C. V., Alonso, M. C., Johnston, P., Phillips, R. G., and Lawrence, P A. (1987) Phenocopies induced with antisense RNA identify the wingless gene. *Cell* **50,** 659–663

60. Giebelhaus, D. H., Eib, D. W , and Moon, R. T. (1988) Antisense RNA inhibits expression of membrane skeleton protein 4.1 during embryonic development of *Xenopus*. *Cell* **53**, 601–615.
61. Bevilacqua, A., Erickson, R. P., and Hieber, V. (1988) Antisense RNA inhibits endogenous gene expression in mouse preimplantation embryos: Lack of double-stranded RNA "melting" activity. *Proc. Natl. Acad. Sci. USA* **85**, 831–835.
62. Bevilacqua, A. and Erickson, R. P. (1989) Use of antisense RNA to help identify a genomic clone for the 5' region of mouse β-glucuronidase. *Biochem. Biophys. Res. Comm.* **160**, 937–941.
63. Bevilacqua, A , Loch-Caruso, R , and Erickson, R. P. (1989) Abnormal development and dye coupling produced by antisense RNA to gap junction protein in mouse preimplantation embryos *Proc. Natl. Acad. Sci. USA* **86**, 5444–5448.
64. Katsuki, M., Sato, M., Kimura, M., Yokoyama, M., Kobuyashi, K., and Nomura, T. (1988) Conversion of normal behavior to *shiverer* by myelin basic protein antisense cDNA in transgenic mice. *Science* **241**, 593–595.
65 Kathory, R., Clapoff, S , Darling, S , Perry, M. D., Moran, L A , and Rossant, J. (1989) Inducible expression of an *hsp68–lacZ* hybrid gene in transgenic mice. *Development* **105**, 707–714.
66. Minty, A., Blau, H., and Kedes, L. (1986) Two-level regulation of cardiac actin gene transcription: Muscle-specific modulating factors can accumulate before gene activation. *Mol Cell. Biol.* **6**, 2137–2148.
67. Russell, E. S. (1979) Hereditary anemias of the mouse: A review for geneticists. *Adv. Genet.* **20**, 357–459.
68. Baumann, N. (1980) *Neurological Mutations Affecting Myelination*. Elsevier/North Holland Biomedical, Amsterdam
69. Sidman, R. L., Green, M. C , and Appel, S. H. (1965) *Catalog of the Neurological Mutants of the Mouse*. Harvard University Press, Cambridge, MA, pp 1–82.
70 Kawamata, J. and Melby, E C , Jr., eds. (1987) *Animal Models: Assessing the Scope of Their Use in Biomedical Research*. Liss, New York, pp. 1–384.
71. Nicholas, F. W. (1987) *Veterinary Genetics*. Clarendon, London, pp. 1–580.
72 Minor, R. R , Wootton, J A. M , Prockop, D. J., and Patterson, D. F (1987) Genetic diseases of connective tissues in animals. *Curr. Prob. Dermatol.* **17**, 199–215.
73. Dodds, W. J. (1988) Third international registry of animal models of thrombosis and hemorrhagic diseases. *ILAR News* **30**, 3–32
74 Wadvoort, H. C. (1983) Glycogen storage diseases in animals and their potential value as models of human disease *J. Inherited Metab. Dis.* **6**, 3–16.
75. LaVail, M. M. (1981) Analysis of neurological mutants with inherited retinal degeneration. *Invest. Opthalmol. Vis Sci* **21**, 638–657.
76. Skow, L. C., Burkhart, B. A , Johnson, F. M , Popp, R. A., Popp. D. M , Goldberg, S. Z., Anderson, W. F., Barnett, L. B , and Lewis, S. E. (1983) A mouse model for β-thalassemia *Cell* **34**, 1043–1052.
77. Peters, J., Andrews, S. J , Loutit, J. F., and Clegg, J. B (1985) A mouse-globin mutant that is an exact model of hemoglobin Rainier in man *Genetics* **110**, 709–721
78. Lewis, S., Erickson, R P , Barnett, L. B., Venta, P. J , and Tashian, R. E (1988) Ethylnitrosourea-induced null mutation at the mouse *Car-2* locus. An animal model for human carbonic anhydrase II deficiency syndrome. *Proc Natl. Acad. Sci. USA* **85**, 1962–1966

CHAPTER 33

Molecular Biology and Medicine

Ethical Implications

Trefor Jenkins

1. Introduction

The recombinant DNA revolution has, since its very birth, been accompanied by many thorny ethical problems; some are new, but most are variations on dilemmas that have confronted scientists and medical practitioners for a very long time.

Concerned about the possible hazards that their newly espoused technology might cause, Singer and Soll (1973) wrote to the president of the National Academy of Sciences and the Institute of Medicine "on behalf of a number of scientists, to communicate a matter of deep concern. ... We presently have the technical ability to join together, covalently, DNA molecules from diverse sources. ... Certain such hybrid molecules may prove hazardous to laboratory workers and to the public. Although no hazard has yet been established, prudence suggests that the potential hazard be seriously considered" *(1).*

The Asilomar conference of February 1975 was the outcome of this concern, and the various worries and apprehensions of the scientists were clearly expressed there—one of the few occasions on which a community of scientists took the initiative to question the consequences of its own research. With hindsight, it is evident that the concerns were exaggerated and the subsequent *NIH Guidelines* (1976) not strictly necessary *(2)*. The scientists had made the mistake of not inviting medical microbiologists to Asilomar, for they were

From: *Methods in Molecular Biology, Vol. 9: Protocols in Human Molecular Genetics*
Edited by· C Mathew Copyright © 1991 The Humana Press Inc., Clifton, NJ

the people who, with their extensive experience in handling dangerous microorganisms, could have reassured the molecular biologists on this score. Representatives of environmental organizations, because of their experience in monitoring noxious agents, also should have been included in the discussions, and experts in occupational health would have had contributions to make as well. In addition, ethicists, theologians, and community leaders might have enriched the debate, but in the past, scientists and physicians have tended to discount their possible contributions. An American medical ethicist, writing on another topic, has somewhat cynically, but no doubt with some justification, commented, "Ethics is generally taken seriously by physicians and scientists, only when it either fosters their agenda or does not interfere with it" *(3)*.

Any discussion of the implications of recombinant DNA technology for the practice of medicine should take place within an ethical framework, buttressed by the three major ethical principles that inform all decision making: beneficence, autonomy, and justice, and the secondary principles that derive from them, namely, confidentiality, truth-telling, and promise keeping *(4)*. Although these principles feature prominently in courses on biomedical ethics offered to medical students and medical practitioners, medical scientists, it is to be hoped, would also perform their work in ways that are guided and informed by these same principles. The concept of the "greater medical profession" proposed some years ago by Sir Theodore Fox, the famous editor of the *Lancet* 1945–1966, would include these scientists as well as other healthcare workers *(5)*. The laboratory scientist who is doing the work to establish whether a patient is a presymptomatic carrier of the gene for Huntington's disease is as much an integral member of the "greater medical profession" as is the physician who is counseling the patient.

Beneficence, or acting in the best interests of the patient, might well necessitate inconvenience to oneself. Respect for the autonomy of the patient will ensure that beneficence does not get out of hand and slide into paternalism, and will guarantee the respecting of confidences, truth-telling, and promise keeping, as well as an acknowledgment that the patient has the right to decide for himself or herself; informed consent will guarantee protection for subjects participating in research projects, including submitting to experimental treatment regimens. It will serve to restrain the over-enthusiastic researcher. Espousing of justice will ensure that patients and research subjects are treated fairly, and the burdens and benefits of the research distributed fairly and evenly; discrimination between people will only be acceptable if the worst off are favored *(6)*. What rules will guide the researcher in making concrete decisions like choice of research topic? Will we tackle problems, the solution of which will benefit the wealthy elite, or will concern for the health of the deprived minorities in the developed countries and of

the inhabitants of the developing world direct us into different fields of research? Will we sell our discoveries to the highest bidder, irrespective of the business ethics of the particular company? Will we show an interest in the uses to which scientific discoveries are put?

Advances in the field of molecular biology are taking place at a breathtaking pace, and it is the application of these discoveries to preventive and predictive medicine, to new reproductive technologies, to enhancement gene therapy and eugenics, and to the human genome initiative that are, perhaps, of most concern at present.

2. Preventive and Predictive Medicine

Molecular biology is making it possible to screen for increasing numbers of genetic diseases. This is not a new concept, because for over a generation most newborns in First World countries have been tested for phenylketonuria (PKU), congenital hypothyroidism, and other inborn errors of metabolism, thereby facilitating early intervention. Carrier detection programs have also been successfully implemented, but to date, these have been directed at only relatively small high-risk populations, e.g., Tay-Sachs disease carriers in the Ashkenazi Jewish people of the USA, Canada, Israel, and so on; the sickle cell trait in ethnic minorities in various countries; and the thalassemias in whole populations in Cyprus and Sardinia, as well as in ethnic minorities in countries like the United Kingdom and Australia.

With the successful cloning of the gene for cystic fibrosis (CF), it is now possible to detect about 70% of heterozygotes for this condition, and further mutations are being identified at such a rate that it can be confidently predicted that, before long, virtually all carriers will be identifiable. In the countries of northwestern Europe and in the countries that have been peopled by emigrants from them, approx 1 person in 20 is a carrier of one of the CF mutations, i.e., 15 million in the USA and 3 million in the United Kingdom, for a start! Screening for these CF carriers has already begun in spite of the reservations of leading individual researchers and the American Society of Human Genetics, meeting in November 1989. A number of biotechnology companies in the USA have marketed their tests, and the cost is apparently about $100 per person; they offer no counseling, but recommend that the referring physicians should do this. Michael Kaback, who pioneered the screening of the United States Jewish community for Tay-Sachs disease carrier status in the early 1970s, believes that better education of the public is an essential prerequisite to whole population screening *(7)*. It is to be hoped that the enthusiasm of molecular biologists (P. Goodfellow, 1989, argued that "carrier-testing is worthwhile now" *(8)*) will be tempered by the experience gained from earlier screening programs and by the real fears that, in the US context,

the results of such tests could be used to deny health insurance to the babies born to "irresponsible" parents who, in spite of knowing that they were carriers, nevertheless took the 1 in 4 chance of having an affected child *(9)*.

The cystic fibrosis gene cloning story is one of the major successes of molecular biology, and was able to be written scholarly and well, because of the collaboration (and competition) between researchers in a number of countries. Prenatal diagnosis (followed by selective abortion of affected fetuses) is already easier and more accurate, and can be done more quickly; the identification of "at risk" couples, i.e., carrier married to carrier, will obviously proceed among the better informed segments of the community; it remains to be seen whether affected infants are going to derive significant benefit from the discovery *(10)*. Research into gene or protein therapy for CF should obviously continue, but support for these approaches "should not allow all the resources to be diverted at the expense of improving existing lines of treatment" *(11)*.

There are a number of dominantly inherited diseases that may become manifest later in life, and because of this fact, they pose some unusual ethical problems. Huntington's disease (HD), with onset in the 4th or 5th decade, is probably the best known, but other neurological disorders like the hereditary ataxias and familial Alzheimer's disease can also exhibit late age of onset. HD has attracted the most attention for a number of reasons, including its high prevalence in Caucasoid populations and the increasing dementia that is an invariable feature of the condition during the patient's last 10 or so years. Although linkage of the disease locus to an anonymous DNA marker was demonstrated in 1983 *(12)*, the gene has still not been cloned. For some time after the demonstration of linkage, Gusella and his colleagues declined to make the probe available to the scientific community at large, stating as their reason that its use was still experimental and that heterogeneity had not been excluded. They were challenged on their stand, and it was claimed that the relatives of patients with HD were being unfairly deprived of a presymptomatic diagnostic test, that Gusella and his colleagues were behaving unethically by withholding information and resources from fellow scientists, and that *Nature* should "neither publish nor refer to work which cannot be validated" *(13)*. Gusella replied that his critics were confusing scientific research with clinical practice and that he was behaving responsibly and in the best interests of patients because of the possible genetic heterogeneity of HD *(14)*. He felt that he would be abrogating his social responsibility if he behaved differently because, as he put it, "A scientist cannot ignore the social consequences of his work, especially in medicine" *(14)*.

It would be difficult to disagree with Gusella when he avers that a scientist cannot ignore the social consequences of his work. The scientist may not, however, be the best-placed person to decide how or if his or her discovery

should be used—"that, as in the case of the [atomic] bomb, is a political decision" *(15)*. It was certainly Gusella's responsibility to show how reliable his probe was for clinical use, and there was also a reasonable expectation by physicians and patients that this should be achieved with a minimum of delay. Whether such action by Gusella and his colleagues retarded progress in the quest for the HD gene is not known.

When the original probes (and other superior ones) for HD were eventually distributed for presymptomatic and prenatal diagnosis, it was done on the understanding that they would be used according to guidelines approved by the supplier, and agreed to by the researcher and a representative of the institution that employed him or her. More recently, a committee consisting of representatives of the International Huntington Association (IHA) and the World Federation of Neurology (WFN) Research Group on Huntington's Disease has drawn up a series of recommendations concerning the use of a predictive test for the early detection of HD; they were approved by the IHA and WFN in June 1989 *(16)*. They seem reasonable and sound, embrace the principles of beneficence and respect for the autonomy of the patient undergoing the predictive test, and should be of help to physicians, counselors, and laboratory scientists. The considerable experience of the members drawing up the document is apparent in almost every recommendation, and their compassion for the individuals at risk is also evident. One of the recommendations is that "each participant should be able to take the test independently of his/her financial means" and is accompanied by the comment: "Each national lay organization should use its influence by advocacy with government departments, public and private health insurers, etc. to reach this goal" *(16)*.

Medical researchers had long been used to sharing information, and even reagents, in an informal way. The exchange of antisera to detect red cell membrane antigen systems (blood groups) characterized much of the early work on the blood groups in the period 1945–1960, and even later. The "commercialization" of these reagents eventually took place, however, and their worldwide distribution is now largely ensured in this way. The demand for DNA probes accelerated very rapidly during the 1980s, and the discoverers of most of them have been relieved that certain institutions have been willing to accept responsibility for their distribution—at a fee. When some obvious commercial application was apparent, patents were applied for and companies set up to exploit the products for financial gain; the discoverer, or the institution for which he or she works, has usually benefited financially. Commercial firms, some set up specially for the purpose, have, in addition, invested large sums of money in recombinant DNA research in the hope of being able to exploit the discoveries made by employee scientists.

When the largest of the biotechnology groups, Genentech, was recently taken over by a giant pharmaceutical company, it was claimed that there had

been disquiet among Genentech's researchers "due in part to a decision by its management to force the scientists to set priorities for their work in response to a worsening of Genentech's financial position" *(17)*. Market forces, it seems, will increasingly determine the direction in which much basic research will proceed in the First World.

3. New Reproductive Strategies

Prenatal sexing and the termination of the male fetus because the mother is a carrier of an X-linked recessive disorder has been accepted practice for a number of years. Terminations were performed because the male fetus had a 1 in 2 risk of being affected with Duchenne Muscular Dystrophy, severe hemophilia A, or X-linked mental retardation, to give a few examples.

Prenatal sexing and termination of the fetus merely because it is not of the desired sex has been carried out in many countries, and one American biomedical ethicist pointed out that the practice may be justified if the particular society sanctioned abortion on demand *(18)*. When asked what they would do if approached by a couple who have four healthy daughters and who now want prenatal sexing with a view to aborting a female fetus, 34% of US genetic counselors said that they would perform the investigation, and another 28% would offer a referral *(19)*. Of the UK geneticists, only 9% would approve and 15% would refer; among Indian geneticists, the figures were 37% and 15%, respectively *(19)*. Whereas American geneticists based their responses on respect for patient autonomy, those from India felt that such action might help to limit the population increase or prevent the suffering or early death of unwanted females *(20)*. Maharastra state enacted strict legislation in an attempt to curb the practice in 1988, but it is alleged that in a recent test case, the government refused to act against a private clinic that advertised sex determination services *(21)*.

The recent vogue for other more elaborate novel reproductive strategies also needs to be questioned, and the practitioners required to justify their research programs. What, for instance, can be the justification for preimplantation genetic diagnosis or embryo selection? Direct biopsy of the blastomere followed by the polymerase chain reaction (PCR) enables the presence of a suspected genetic disorder to be established and the affected embryo to be discarded; if the disease is not present, the embryo can be implanted. Likewise, an ovum removed from a heterozygous carrier of a specific disease can be shown to carry the mutant gene or its normal allele by PCR amplification and analysis of the first polar body, which, if found to contain the mutant gene, would mean that the other oocyte contains the normal allele and vice versa, i.e., the genotypes will be reciprocal *(22)*. In the case of polar body analysis, the patients and the practitioner implacably opposed to abortion

might be accepting of the technique; the blastomere biopsy and the discarding of the embryo when a positive diagnosis is obtained would be considered an abortion by the "hardliners." In response to the argument that psychological stress is significantly less with these approaches than it is with chorionic villus sampling (CVS) at about 8–9 wk gestation, we need to be reminded that the success rate of in vitro fertilization (IVF) is, in the best of centers, less than 20%, with three attempts at embryo transfer. It is common practice in IVF programs to produce an excess of embryos for implantation, and this usually necessitates the disposal of those not used—a practice that would be unacceptable to people disapproving of even very early abortion. IVF has proved its usefulness for certain well-defined causes of infertility; its use in association with preimplantation genetic diagnosis would seem to be an extremely expensive (and even extravagant) technology when alternative strategies like amniocentesis and CVS are deemed acceptable by the vast majority of couples at risk for a child with a single gene disorder that is amenable to prenatal diagnosis.

4. Gene Therapy

Some researchers will obviously disagree with my assessment of reproductive strategies, and one has stated that "it would be slightly blinkered to confine the discussion (on the prevention of genetic disease) to prenatal diagnosis. We are all heading for primary prevention. Most molecular geneticists want to cure and treat disease, not do abortions" *(23)*. Primary prevention would include gene therapy on the zygote, and because this would entail an alteration of the germ cell line, researchers have not yet attempted it on humans, although it is technically feasible. Before the introduction of germ line gene therapy, however, somatic cell gene therapy in humans should have been shown to be effective and safe; adequate animal studies must have been done using the same vectors and techniques that are contemplated for use in humans, and the inserted DNA must be shown to be "expressed in the appropriate tissues and at the appropriate times"; and "there should be public awareness and approval of the procedure" *(24)*. Because germ line gene therapy is a very different and unique form of treatment that will affect future generations, its introduction should rightly have the prior approval of an informed public. The human gene pool is at risk in such an enterprise, and since the human gene pool is the possession of all of humankind, international approval would be desirable before its introduction. Before being able to give such approval, the public should have a good understanding of the implications of germ line gene therapy.

Somatic cell gene therapy for the treatment of severe disease is considered ethical—merely an extension of other modalities of therapy—because

it can be supported by the moral principle of beneficence. There has, nevertheless, been much controversy over it, and the first clinical application was approved by the National Institutes of Health and the Food and Drug Administrations as recently as January 1989, "the most thorough prior review of any clinical protocol in history. It was approved only after being reviewed fifteen times by seven different regulatory bodies. ... (it) demonstrates that the concept of gene therapy raises serious concerns" *(25)*.

If somatic cell gene therapy is capable of curing severe genetic disease, can it also be used to enhance certain "normal" characteristics? Would it be permissible to insert a gene in order to "enhance" the production of growth hormone in an infant, thereby producing a person of extremely high stature—a champion basketball player, say, or to enhance memory or intelligence? Such a procedure might well disturb a delicate balance within or between cells of the body, thereby adversely affecting essential biochemical pathways. Although enhancement gene therapy may not be acceptable for such frivolous purposes, it is not difficult to imagine situations in which it might be justifiable as a strategy in preventive medicine. For example, individuals with familial hypercholesterolemia have insufficient or defective receptors for LDL cholesterol on their cells. As a result, they produce excessive amounts of endogenous cholesterol and are unable to clear the substance from their blood. Their cholesterol levels remain high, and increase their risk of ischemic heart disease, which may prove fatal in young adulthood. A gene for normal LDL receptor production inserted into a patient's genome in early life might well enhance receptor production and protect him or her from suffering from myocardial infarction in the third or fourth decade of life.

5. The New Eugenics

Yet another level of "gene therapy" can be considered. Because it would attempt to "improve" the normal genetic constitution of an individual, influencing personality, character, fertility, and intelligence, as well as physical, mental, and emotional characteristics, it is referred to as "eugenic" genetic engineering. Such technology, even if it is still futuristic because, as yet, we know so little about the inherited components of these traits, fascinates the readers of science fiction and even the popular press. Eugenics was discredited immediately before and during the Second World War—largely because of the perverted excesses of Nazi physicians and scientists in programs of, what they called, race hygiene. Although human genetics as a science has developed enormously since the Second World War, we are still so ignorant of the genetic basis of the "socially interesting and useful" traits, some of which are listed above, that it would be extremely unwise to meddle with them unless it were for definite therapeutic reasons.

In modern times, the United States and Germany have both passed through periods when power was abused to further eugenic goals *(26,27)*. Constant vigilance is needed if we are to resist drifting into a new eugenic age. Small "improvements" might constitute the "thin edge of the wedge," and once begun, it might be impossible to know where to draw the line. "Therefore, gene transfer should be used only for the treatment of serious disease and not for putative improvements" *(26)*. The dangers of the abuse of recombinant DNA technology to manipulate the genome of human beings deliberately to serve perverted sociological ends can best be guarded against by a well-informed public, committed to upholding the highest values. Scientists have an important duty to contribute to informing the public, and if anyone doubts the capability of the public to make the correct decision, he or she may take comfort and derive encouragement from the conviction of Thomas Jefferson: "I know no safe depository of the ultimate powers of the society but the people themselves; and if we think them not enlightened enough to exercise that control with a wholesome discretion, the remedy is not to take it from them, but to inform their discretion by education" *(28)*.

6. The Human Genome Project

The methodology already exists for the construction of a linkage and a physical map of the human genome, the ultimate physical map being, of course, its complete nucleotide sequence. The latter would be a mammoth undertaking by any standards, and the mere listing in order of every one of the genome's 3×10^9 bp would, it has been estimated, fill a million-page book. Thousands of person-hours have already been spent discussing the pros and cons of such a project, and the ordinary person in the street might be forgiven for thinking that the records of the debates about the project if collected together would already fill a very large book, perhaps not quite a million pages long!

One of the most enthusiastic protagonists of the sequencing idea has been James Watson, the director of the Center for Human Genome Research of the US National Institute of Health. Watson was successful in convincing the US Congress to allocate substantial funds for the project (they will soon be $200 million per year with a commitment to $3 billion spread over 15 years), but in February 1990, he reluctantly acknowledged that it would be at least five years before a dedicated effort to sequence the human genome could be launched, owing to the unacceptably high costs involved *(29)*. Until automated sequencing technology is developed to the level at which the costs are considerably reduced, the project cannot begin. Commercial companies in the US and elsewhere (particularly in Japan) are obviously very interested in developing the technology and distributing it worldwide.

The forum at which Watson spoke was an international meeting organized by UNESCO (United Nations Educational, Scientific, and Cultural Organization) and the Scientific Committee on Genetic Experimentation (COGENE) of the International Council of Scientific Unions. It was appropriate, therefore, that he should also use the occasion to soften his previous position, namely, that only those countries that provided financial contributions to the sequencing project should be allowed access to its results. Now, he felt that the results should be held back from immediate publication, but only until they had been "fully interpreted" *(29)*. Watson has been accused of wanting to keep secret the results of US-sponsored genome research "lest the Japanese use them to develop marketable biotechnology products" *(30)*.

In an attempt to assess the potential impact of the human genome project on the detection, diagnosis, prevention, and treatment of human disease, Friedmann concluded that "reverse genetics," which is conceptually relatively straightforward, will not be the means of identifying all the disease loci because of the enormous technical difficulties surrounding such an approach *(31)*. Although a detailed genetic linkage map of the human genome will provide linked RFLPs for the eventual isolation of every disease locus, it will not provide information on intergenic and regulatory regions that may contribute to the expression of diseases. The physical map will provide the means of solving these kinds of problems.

The cloning and characterization of disease-causing genes have usually resulted in improved diagnosis, screening, and even prevention; to date, there has been little to report by way of improved therapy. One needs merely to think of the hemoglobinopathies to realize how little we can offer. Screening programs for sickle cell anemia, and for carriers of the gene, in the United States in the 1970s were not unqualified successes, owing largely to inadequate public education; and the stigmatization of carriers identified by these programs was a serious undesirable side effect. Nevertheless, it seems more than likely that screening programs for genetic disorders will soon be introduced and justified by cost-efficiency considerations. Cystic fibrosis might be one such disease in Western European and North American countries, whereas the thalassemia syndromes in, for example, Cyprus, Sardinia, and Thailand are others *(32)*.

Many of the ethical issues raised by the human genome project have taxed the minds of thoughtful medical scientists over recent years. Presymptomatic diagnosis of serious progressive diseases for which there are no effective therapies (Huntington's disease is the best-known example) places in the hands of patients information that they may not be capable of handling. Prenatal diagnosis is likely to become an ever-increasing service demand, thereby increasing the rift between antiabortion pressure groups and those of a prochoice persuasion. When RFLP gene tracking has to be pur-

sued at the family level, problems concerned with confidentiality and access to useful information will become more acute; the inadvertent discovery of exclusions of paternity may also be unwelcome consequences. Employers and insurance companies may demand access to confidential genetic information before employing someone or accepting him or her for medical or life insurance—discriminatory practices that may be difficult to prevent. Legislation may be required to prevent these practices and the use of such information as a tool for injustice and equity; perhaps the law will restrict the accumulation of such genetic knowledge.

It is apparent that many other ethical and societal issues are, and will increasingly be, raised by the human genome project, so the allocation of 1–3% of the US genome project budget to research these issues is to be welcomed. The ethical issues are also a major concern of HUGO, the international body set up by the scientific community to coordinate worldwide efforts to sequence the human genome *(33)*.

Third World scientists may be forgiven for being skeptical about the claims of colleagues like Watson when, referring to the human genome project, he says, "I see an extraordinary potential for human betterment ahead of us. The time to act is now" and "How can we not do it? We used to think our fate was in our stars. Now we know, in large measure, our fate is in our genes" *(34)*. The main need of the developing countries is the rapid application of research findings that have already been made; the problem is one of distribution and implementation. In addition, research needs to be directed into finding vaccines for the major infectious diseases that are not problems in the First World. It is hard to believe that, if malaria were as prevalent in the First World as it is in the Third World, we would not by now have effective vaccines to deal with the problem. Nevertheless, the developing countries can be assisted by grand schemes, like the Human Genome Project. Scientists from the developing countries can, by participating in the research, acquire skills and expertise in recombinant DNA technology *(35)*, but if this takes place in foreign countries, incentives must be provided for them to return to their home country where research into locally relevant problems may be initiated. The present situation, in which more than one-half of Third World postgraduates who go to US universities stay on in that country, is totally unacceptable. The International Centre for Theoretical Physics at Trieste in northern Italy, founded by Abdus Salam, with a major goal of training young Third World scientists in an atmosphere that encourages their return to their home countries, could serve as a model on which centers devoted to biotechnology and the like could be founded *(36)*.

The prospect of sequencing the human genome excites molecular biologists, and its successful completion would be another milestone on the highway of humankind's intellectual progress and achievement. In the proc-

ess, genes responsible for many diseases would be identified, as would many that predispose to ill health. Numbers of such genes have already been identified by individual researchers working in an uncoordinated way, and more can be expected to be found in this piecemeal fashion. The glamour of a megaproject has, however, captured the imagination of scientists, as well as some governments, and is likely to proceed. The important ethical issue is that of distributive justice, or the allocation of scarce resources. Should $200 million per year be spent by the US NIH on a single project directed by a small number of scientist/administrators when there already exists a satisfactory well-tried system of funding research? It is claimed that funds will not be diverted from other biological research, and $200 million per year is not all that large if one recalls that the Cystic Fibrosis Foundation spent $120 million in four years on finding the gene for a single disease, and, in addition, many millions were also spent on it by other foundations and also government agencies *(37)*.

The very eminent British scientist, Sydney Brenner, argues that it would be wasteful to sequence the whole genome now. Because 98% of the genome is "junk," Brenner thinks it is necessary to sequence only the other 2% of it *(38)*. Those who disagree claim that the only way to prove that the junk DNA is unimportant is to sequence it now and not leave it for a future generation of scientists to tackle, as Brenner suggests.

In urging scientists to proceed with the Human Genome Project, Koshland has reminded us of "the immorality of omission—the failure to apply a great new technology to aid the poor, the infirm, and the underprivileged" *(37)*. An American physician responded, pointing out that such an admonition "might sound cynical to healthcare workers in Third World countries who deal with countless children not even vaccinated against polio or tetanus or to physicians and nurses in our own country unable to apply state-of-the-art medicine to 'medically indigent' people whose care appears to be of little concern to the rest of society. In these cases, the 'great' technology already exists, but politics and economics prevent its application" *(39)*. Koshland also believes that "the potential risks from the new knowledge gained by sequencing the human genome appear on close examination to be old problems revisited." The implication is that we have already dealt with these problems in a satisfactory manner, but if some of these problems are mandatory screening, confidentiality, privacy, and discrimination, it is debatable whether this is, in fact, the case.

The development of these exciting new technologies is very likely to lead to screening on a scale, many orders of magnitude greater than has been the case to date; the goal of such screening will be to alert individuals to their status and to encourage them to mate with noncarriers, or to use artificial insemination or other reproductive strategies. Our first, admittedly lim-

ited, attempts at screening have not been uniformly successful, and have been accompanied by many undesirable side effects. Holtzman (1989) has recently considered this subject from an historical perspective and enthusiastic, young molecular biologists who see screening as the answer to most medical genetics problems would do well to read this extremely well-written book *(40)*. Imposing genetic tests on people against their wishes constitutes a new eugenics, with motives not very different from those of the early eugenicists. Of course, we know more genetics than they did, and we have at our disposal greatly improved techniques to facilitate accurate screening. However, on the subject of whether society has the right to enforce such testing, Holtzman writes:

> When fully educated and informed, most people will probably accept carrier testing, prenatal diagnosis, and the abortion of fetuses who are destined to develop severe disease in infancy or childhood. Regardless of how many refuse testing, society's imposing of its will on them may exact a greater price in tearing our social fabric than would caring for their affected offspring (*40*, p. 229).

Another level of concern raised by the Human Genome Project "relates to the fact that powerful technologies do not just change what human beings can do, they change the very way we think—especially about ourselves" *(3)*. Potential parents might resort to complete screening of embryos and only implant those that are considered to be "high-grade." In addition to putting a specific price on human characteristics, an attitude could develop that would see children as commodities, existing to satisfy the demands of parents and even societies, without regard for the children's own rights and interests. Such a program has been viewed by one eminent scientist as an attempt to "perfect" human individuals by "correcting" their genomes in conformity, perhaps, to an ideal, "white, Judeo-Christian, economically successful genotype" *(41)*.

7. The Motives and the Social Responsibility of Scientists

J. Robert Oppenheimer, speaking to a group of scientists in 1945, soon after stepping down from the Los Alamos directorship, addressed the basic question of *why* scientists had built the atomic bomb. There were motives like fear that Nazi Germany would build it first, and that it would shorten the war, but he believed that the basic motivation was inherent in the very nature of science itself, and said:

> If you are a scientist you believe that it is good to find out how the world works; that it is good to find out what the realities are; that it is good to turn over to mankind at large the greatest possible power to

control the world and to deal with it according to its lights and its values. ... It is not possible to be a scientist unless you believe that the knowledge of the world, and the power which this gives, is a thing which is of intrinsic value to humanity, and that you are using it to help in the spread of knowledge, and are willing to take the consequences. (*42*, p. 761).

Do the scientists have a special responsibility for the consequences of their discoveries? Some famous scientists have argued that they do, and some have even refused to give information to enquirers for fear that it would be misused. Oppenheimer argued, however, that the suggestion that a scientist should "assume responsibility for the fruit of his work ... appears little more than an exhortation to the man of learning to be properly uncomfortable" (*43*, p. 67) and he was supported in this view by other leading physicists of the day. The dissenting view was expressed by Bertrand Russell:

> The scientist is also a citizen; and citizens who have any special skill have a public duty to see, as far as they can, that their skill is utilized in accordance with the public interest. ... It is impossible in the modern world for a man of science to say with any honesty, "My business is to provide knowledge, and what use is made of the knowledge is not my responsibility" (*44*, p. 391).

The big issue of technology that engaged these great minds was, of course, nuclear energy, in which there was no uncertainty about the effects of its use and misuse. In the case of recombinant DNA technology, there is still uncertainty about the effects, and as a result, social guidelines, rules, laws, and ethical codes for its use have not yet been clearly and unambiguously formulated. It is not, therefore, easy (or perhaps compelling) to call for scientists to "be responsible" in the rapidly expanding field. One valuable piece of advice has, nevertheless, been given by Arthur Galston, the scientist whose research on plant maturation factors led to the development and military use of the defoliant, Agent Orange (*45*). Having originally thought that one could avoid involvement in the antisocial consequences of science by carefully choosing one's field of research, he eventually came to realize that almost any scientific finding can be put to immoral use. He wrote:

> ...the only recourse for the scientist concerned about the social consequences of his work is to remain involved with it to the end. His responsibility to society does not cease with publication of a definitive scientific paper. Rather, if his discovery is translated into some impact on the world outside the laboratory he will, in most instances, want to follow through to see that it is used for constructive rather than anti-human purposes (*45*).

We all know, however, that some scientists do not feel any moral imperative to get involved in such activities at the society level, and Galston feels that they might "shun such activity, either through timidity, aversion to political argumentation, or a feeling that others, better trained, should handle such problems" *(45)*. The time has probably passed when individual scientists can do much about these problems—as individuals. The novel and diffuse ethical problems confronting scientists can be most effectively handled by collective approaches based on a sense of stewardship and professional responsibility. The days of heroic initiative and sacrifice epitomized by Rudolph Virchow *(46)* are probably gone forever, but this is not to say that such visionaries do not have a major galvanizing role to play within the communities of scientists to be found in many countries around the world.

Erwin Chargaff (1987) eloquently expressed his fears about the adverse effects that in vitro fertilization and the freezing of embryos might have on the person "produced" in this way; about the newly acquired ability to modify the hereditary apparatus; about the "manufacture of human embryos with the only purpose of crushing and extracting them" *(47)*. He is of the opinion that excesses of scientific research are taking place and calls for scientists to exercise restraint in asking certain questions, a sacrifice "that even the scientist ought to be willing to make to human dignity." If Csikszentmihalyi (1985) is correct, and enjoyment (mediated presumably by endorphin-like substances acting on the brain) in doing certain tasks is a powerful force, capable of driving scientists toward creative efforts, then this must be recognized as such and the force tamed or directed *(48)*. One famous scientist has admitted, "Little by little, I became addicted. Doing experiments turned into a mania, a drug I could not do without" *(49)*. It is, perhaps, necessary to ask how much molecular biological research, or any scientific research, for that matter, is inspired by altruistic motives with the goal of improving the lot of humankind, and how much by fascination with the problem itself.

The dangers of governmental interference with research have, however, been illustrated by recent events in the United States. Transplantation of fetal brain cells into patients suffering from Parkinson's disease is considered to be a promising line of research and is being pursued in Sweden, Canada, and the UK. It has, however, been virtually banned in the US since the secretary of health and human services announced toward the end of 1989 that no federal support would be given for research involving tissue from elective abortions. In the same way, research using "surplus" embryos produced in IVF programs is also effectively banned. Scientists are worried that these steps, taken in response to the lobbying of antiabortion groups, will seriously impair research into human fertility and the etiology of fetal abnormalities. This type of research will continue in other countries, but John Fletcher, a leading American bioethicist, has cautioned that "such ideologically motivated screen-

ing of research recalls the influence of Trofim Lysenko on biological research in the Soviet Union" under Stalin *(50)*.

Müller-Hill, who has researched the activities of scientists during the Nazi era, pleads that scientific truth and cost efficiency should not be regarded as more important than human dignity *(51)*. He cautions that the vast amount of data generated by the Human Genome Project could be used to the disadvantage of individuals with respect to employment or insurance coverage, thereby compromising the dignity of the person. The motive of the German scientists to participate in the "race hygiene" programs was the improvement of the human species by purging the gene pool of putatively inferior elements. This might be considered a laudable motive, but we now realize that they were operating in a society that had sacrificed so many of the values considered fundamental to a civilized community, respect for the dignity and worth of every individual being probably the most fundamental.

8. Conclusion

This review began by recalling that, in the early 1970s, a number of molecular biologists demonstrated their concern for what they perceived to be the hazards that might emanate from their discoveries. The Asilomar Conference and the *NIH Guidelines* resulted from this concern. Some commentators have, with the wisdom of hindsight, been critical of their action *(52)*, but others have commended the scientists for manifesting responsibility and "exemplary stewardship" *(53)*. It is to be hoped that molecular biologists, and indeed all scientists, will adopt an equally responsible attitude toward research findings and technological advances, as well as toward the choice of research project in the first place.

Recombinant DNA technology will have important applications in the field of healthcare: population screening will be instituted on an increasingly large scale, alerting couples to their risks of having a child with serious inherited disorders. Prenatal diagnosis using the cloned gene or synthesized oligonucleotide probes will be feasible, and parents will be given the option of selective abortion of the affected fetus. Presymptomatic diagnosis of diseases like Huntington's disease, myotonic dystrophy, and others will permit individuals to make responsible decisions concerning reproduction. Individuals at genetic risk for work-related damage to their health may opt for work in other fields, always assuming that such jobs are readily available within the society offering the screening tests. Information about an individual's genetic constitution could be misused by employers, insurance companies, governmental agencies, and mischievous people intent on blackmailing individuals with threats of exposing "sensitive" information. Care in ensuring confidentiality and access to these data will need to be exerted by the defined custodians; legally guaranteed safeguards may well be essential.

New reproductive strategies have become feasible because of the rapidly developing technologies that permit preimplantation diagnosis of genetic disease in the few-day-old embryo or in individual human oocytes—now that PCR amplification is available. Because these techniques require IVF, it is unlikely that they will soon replace the well-established method of diagnosis of such diseases by amniocentesis carried out at 14–16 wk gestation or CVS, either transcervically or transabdominally, at about 9–10 wk.

It is ironic that the very common single gene disorders, like sickle cell anemia and the thalassemia syndromes, occur at their highest frequencies in the tropical, developing countries of Africa, India, and South East Asia (as well as in the other countries into which people from these tropical countries have immigrated), countries that, traditionally, have been said to lack the scientific infrastructure necessary to institute preventative programs. As development proceeds, and as infant mortality rates drop to more acceptable figures, say 30–50/1000 live births, the financial and manpower burdens placed on these societies as they try to treat these common, chronic, genetic disorders, will be so great that the healthcare services will be unable to cope and will collapse under the strain, unless carrier detection programs together with prenatal diagnosis and selective abortion are introduced.

Gene therapy at the level of the somatic cell is not yet readily available, and even when it is feasible, it will probably be suitable for only a small number of individuals with generally rare disorders. Germ line gene therapy, successfully performed in experimental animals but often with distressing side-effects, is considered ethically undesirable by most researchers. Enhancement gene therapy may well be demanded by parents, for example, for their child who is of short stature even though his or her height is within the normal range. Some researchers, encouraged by parents, may be tempted to strive for eugenic goals with the very real risk that the dignity and the autonomy of the "manipulated" individual will be seriously compromised. In no field of applied molecular biology will it be more essential to ensure that all work is carried out within a well-developed framework of values, discussed and debated by scientists, theologians, lawyers, and ethicists. A well-informed public will ensure that satisfactory guidelines are laid down and followed.

One of the most exciting and challenging enterprises of modern research in the field of human molecular genetics has been labeled "The Human Genome Project." The prospect of knowing where every structural gene locus is situated in the genome has excited many scientists and in some countries, the politicians have generously responded with funds to permit this megaproject to proceed. Most are agreed on the worthwhileness of the project, although there is difference of opinion on how to set about doing it. The short-term benefits of this type of research are not clearly apparent, even for the First World. It has questionable relevance for the Third World, where

there has been a failure to introduce the already available technology to improve food production, combat infectious disease, control the population explosion, and reduce the incidence of genetic hemolytic anemias.

It is still early days, with recombinant DNA technology having been applied in the field of medicine for barely a decade. Although cynics might argue that our expectations for the new genetics have been naive and unrealistic, others will agree with the evaluation and forecast made by Joshua Lederberg in 1979:

> DNA splicing research, far from being an idle scientific toy or the basis for expensive and specialized aid to the privileged few, promises some of the most pervasive benefits for the public health since the discovery and promulgation of antibiotics *(54)*.

The discoveries resulting from the use of recombinant DNA technology have been phenomenal, and the application of the technology to combat a host of genetic disorders in the better educated families in First World countries ("the privileged few") has been very impressive. The fact that such little impact has so far been made on the public health, particularly in the Third World, is not because of any deficiency of the technology itself. It is caused, rather, by the lack of will and commitment to direct the research effort to combating the major infectious diseases of these countries, to developing engineered food plants suitable for adverse climatic conditions, and to developing new chemotherapeutic agents based on rational structural analysis of the target molecules. The AIDS epidemic, affecting both the Third and First Worlds, and now encroaching on the Second World, should serve as a reminder to scientists in the First World that they ought, if only for motives of self-interest, to direct more of their research efforts into solving the problems of tropical diseases. Many will concur with Lederberg, who sees the AIDS epidemic as a reason for us to "take pride at our sense of fraternity with mankind," and as presenting us with a problem area "in which self-interest and humanitarian interest converge absolutely and certainly" *(55)*. Whereas medical scientists must obviously continue to explore the genetic basis of disease, they must also take more of an interest in the ways in which their discoveries are applied to improve the health of the Third World, where three-quarters of humankind live and where 96% of child and infant deaths occur.

References

1. Singer, M. and Soll, D. (1973) Guidelines for DNA hybrid molecules (letter to the editor). *Science* 181, 1114.
2. NIH Guidelines for research involving recombinant DNA molecules. (1976) *Federal Register* 41, 27, 902–927, 943.

3. Annas, G. J. (1989) Who's afraid of the human genome? *Hastings Center Report* **19(4)**, 19–21
4. Pellegrino, E. D. (1987) The anatomy of clinical ethical judgements in perinatology and neonatology: A substantive and procedural framework. *Semin. Perinatol.* **11**, 202–209.
5. Fox, T. F. (1956) The greater medical profession. *Lancet* **ii**, 779,780.
6. Rawls, J (1971) *Theory of Justice.* Harvard University Press, Cambridge, MA.
7. Joyce, C. (1990) Gene test for cystic fibrosis sparks off screening debate. *New Scientist,* **10 February**, 22.
8. Goodfellow, P N. (1989) Steady steps lead to the gene. *Nature* **341**, 102,103.
9 Colten, H. R. (1990) Screening for cystic fibrosis: Public policy and personal choices. *N. Engl J. Med.* **322**, 328,329
10 Koshland, D E., Jr. (1989) The cystic fibrosis gene story. *Science* **245**, 1029.
11. Knight, R A. and Hodson, M E (1990) Identification of the cystic fibrosis gene. *Br. Med J.* **300**, 345,346.
12 Gusella, J F., Wexler, N S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P C , Ottina, K., Wallace, M. R., Sakaguchi, A. Y., Young, A. B , Shoulson, I., Bonilla, E., and Martin, J. B (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234–238.
13 Watt, D. C , Lindenbaum, R. H , Jonasson, J. A., and Edwards, J. H. (1986) Probes in Huntington's chorea. *Nature* **320**, 21.
14 Gusella, J. F. (1986) Probes in Huntington's chorea. *Nature* **320**, 21,22.
15. Wolpert, L. (1989) The social responsibility of scientists. moonshine and morals. *Br. Med J.* **298**, 941–943
16. IHA and WFN (1990) International Huntington Association and World Federation of Neurology Research Group on Huntington's Disease. Ethical issues policy statement on Huntington's disease molecular genetics predictive test. *J Med. Genet.* **27**, 34–38.
17. Watts, S (1990) Drugs giant takes over Genentech. *New Scientist,* **10 February**, 23.
18. Fletcher, J. C. (1979) Ethics and amniocentesis for fetal sex identification. *N. Engl. J. Med.* **301**, 550–553
19 Wertz, D. C. and Fletcher, J C (1989) Ethics and genetics· An international survey *Hastings Center Report* **19**, No. 4, Special Supplement, 20–24.
20 Wertz, D. C. and Fletcher, J. C (1989) Fatal knowledge? Prenatal diagnosis and sex selection. *Hastings Center Report* **19**, No. 3, 21–27.
21 Rao, R (1990) Sex selection continues in Maharastra. *Nature* **343**, 497
22. Monk, M and Holding, C. (1990) Amplification of a β-haemoglobin sequence in individual human oocytes and polar bodies. *Lancet* **335**, 985–988.
23 Bell, J. (1989) Using human genetic information. (Report on Ciba Foundation 40th Anniversary Symposium, Berne, Switzerland). *Lancet* **ii**, 58,59.
24 Anderson, W. F. (1985) Human gene therapy· Scientific and ethical considerations *J. Med Philos.* **10**, 275–291
25. Anderson, W F. (1990) Genetics and human malleability. *Hastings Center Report* **20**, No. 1, 21–24
26. Kevles, D. J (1985) *In the Name of Eugenics.* Alfred A. Knopf, NY.
27 Müller-Hill, B. (1988) *Murderous Science* (G. R Fraser, trans.), Oxford University Press, Oxford, UK.
28. Padover, S. K. (ed) (1939) *Thomas Jefferson on Democracy.* Mentor Books, The New American Library, New York, pp. 89,90.

29. Hughes, S. (1990) Five-year wait predicted for genome project. *New Scientist* **10 February,** 23.
30. Randal, J. (1989) The human genome project. *Lancet* **ii,** 1535,1536.
31. Friedmann, T. (1990) The human genome project—some implications of extensive "reverse genetic" medicine. *Am. J. Hum. Genet.* **46,** 407–414
32. WHO (1982) Community control of hereditary anaemias: Memorandum from a WHO meeting. *Bull. WHO* **61,** 63–80.
33 McKusick, V. A. (1989) Mapping and sequencing the human genome. *N. Engl. J. Med.* **320,** 910–915.
34. Jaroff, L. (1989) The gene hunt. *Time,* **20 March,** 62–67
35 Grisolia, S. (1988) Mapping the human genome. *Hastings Center Report* **19,** No. 4, Supplement 18,19.
36. Hall, N. (1990) A unifying force for Third World science *New Scientist,* **27 January,** 31,32.
37. Koshland, D. E. (1989) Sequences and consequences of the human genome *Science* **246,** 189.
38. Brenner, S. (1990) The human genome: the nature of the enterprise. *Human Genetic Information: Science, Law and Ethics.* (Ciba Foundation Symposium, 149.) Wiley, Chichester, pp. 6–17.
39. Cooper, D. M. (1989) Human genome programme. *Science* **246,** 873,874.
40. Holtzman, N. A. (1989) *Proceed with Caution. Predicting Genetic Risks in the Recombinant DNA Era.* The Johns Hopkins University Press, Baltimore, MD and London, UK.
41. Luria, S. E. (1989) Human genome programme. *Science* **246,** 873
42. Rhodes, R. (1988) *The Making of the Atomic Bomb.* Penguin Books, London, UK.
43. Oppenheimer, J R. (1948) Physics in the contemporary world. *Bulletin of the Atomic Scientists* **4,** 65–68, quoted by Lowrance (1985).
44. Russell, B. (1960) The social responsibilities of scientists *Science* **131,** 391,392.
45. Galston, A. W. (1972) Science and social responsibility· A case history. *Ann. NY Acad. Sci.* **196,** 223–235.
46. Eisenberg, L. (1986) Rudolf Virchow: The physician as politician. *Medicine and War* **2,** 243–250.
47. Chargaff, E. (1987) Engineering a molecular nightmare. *Nature* **327,** 199,200.
48. Csikszentmihalyi, M. (1985) Reflections on enjoyment. *Perspect. Biol. Med* **28,** 489–497.
49. Jacob, F. (1988) *The Statue Within. An Autobiography.* Basic Books, NY.
50. Beardsley, T. (1990) Aborted research. *Scientific American* **262,** 10.
51. Müller-Hill, B. (1989) Eugenics: The science and religion of the Nazis. Paper presented at Conference on "The Meaning of the Holocaust for Bioethics," Minneapolis, May 17–19, 1989.
52. Watson, J. D. (1981) *The DNA Story: A Documentary History of Gene Cloning* (Watson, J. D and Tooze, J., eds.) W. H Freeman, San Francisco, CA.
53 Lowrance, W W. (1986) *Modern Science and Human Values.* Oxford University Press, Oxford, UK.
54. Lederberg, J. (1979) DNA splicing: Will fear rob us of its benefits? in *The Recombinant DNA Debate* (Jackson, D. A. and Stitch, S. P , eds.), Prentice-Hall, Englewood Cliffs, NJ, pp. 173–180.
55. Lederberg, J. (1989) Introduction: Biomedical science and the third world: Under the volcano (Bloom, B. R. and Cerami, A., eds.), *Ann. NY Acad. Sci.* **569,** xix,xx.

Appendix

Manuals and Data Bases

1. Useful Laboratory Manuals for Molecular Biologists

1. *Nucleic Acids* (Methods in Molecular Biology, vol. 2), edited by J. M. Walker, Humana Press, Clifton, NJ, 1984. This volume contains 53 chapters describing commonly used protocols.
2. *New Nucleic Acid Techniques* (Methods in Molecular Biology, vol 4), edited by J. M. Walker, Humana Press, Clifton, NJ, 1988. A further 44 chapters of molecular protocols.
3. *Guide to Molecular Cloning Techniques* (Methods in Enzymology, vol. 152), edited by S. L. Berger and A. R. Kimmel, Academic Press, New York, 1987.
4. *Molecular Cloning, A Laboratory Manual* (vols. 1-3), by J. Sambrook, E. F. Fritsch, and T. Maniatis, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

2. Human Genetics Data Bases

1. Genome Data Base (GDB). This is a very useful data base for practising Human Molecular Geneticists. It contains information on the chromosomal location of genes, details of all DNA polymorphisms that have been described, and literature references. To register as a user, contact: GDB/OMIM User Support, William H.Welch Medical Library, 1830 E. Monument Street, Third Floor, Baltimore, MD 21205, USA. Telephone: (301) 955-7058 Telefax: (301) 955-0054.
2. On-line Mendelian Inheritance in Man (OMIM).This data base is an on-line version of the full text of Dr Victor McKusick's *Mendelian Inheritance in Man,* which gives details of all inherited disorders that have been described and their chromosomal location, if known. This data base is linked to GDB, and access is provided via the Welch Medical Library (see GDB for address).
3. Data bases of human DNA sequences have been established at the National Institutes of Health, USA (GENBANK) and at the European Molecular Biology Laboratory, Heidelberg, FRG (EMBL). The University

of Wisconsin Genetics Computer Group package (GCG) is a set of sequence analysis programs for use with these data bases. A detailed discussion of the software available and its application can be found in *Nucleic Acid and Protein Sequence Analysis: A Practical Approach*, edited by M. J. Bishop and C. J. Rawlings, IRL Press, Oxford/Washington DC, 1987.

3. DNA Probe Banks

1. ATCC/NIH Repository of Human and Mouse DNA probes and Libraries. Contact: American Type Culture Collection, 12301 Parklawn Drive, Rockville, MD 20852-1776, USA. Telephone (301) 881-2600 Telefax (301) 770-2587.
2. Collaborative Research Incorporated. This is a commercial operation. Contact: Collaborative Research Inc., Biomedical Products Division, 2 Oak Park, Bedford, MA 01730, USA. Telephone (617) 275-0004.
3. UK Human Genome Mapping Project. Contact: HGMP Resource Centre, Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, UK. Telephone (081) 869 3446 Telefax: (081) 869 3807. Cost: £40 (UK) per probe.